

# Perspectivas filosófico-políticas contemporáneas

*Reflexiones y debates para el siglo XXI*



Santiago Ullauri Betancourt

Diego Ortiz-Jaramillo

Nicolás Fuentes Valdebenito

Felipe Álvarez

**Coordinadores**

 **LHE**ediciones

# **Perspectivas filosófico-políticas contemporáneas: reflexiones y debates para el siglo XXI**

Santiago Ullauri Betancourt

Diego Ortiz-Jaramillo

Nicolás Fuentes Valdebenito

Felipe Álvarez

***Coordinadores***

**Título:** Perspectivas filosófico-políticas contemporáneas: reflexiones y debates para el siglo XXI

**Coordinadores:** Santiago Ullaauri Betancourt, Diego Ortiz-Jaramillo, Nicolás Fuentes Valdebenito y Felipe Álvarez

**Primera edición**

**ISBN:** 978-9942-752-38-3

**DOI:**<https://doi.org/10.31207/uhediciones23>

Publicado en octubre de 2026

© **UHEdiciones** • Universidad Hemisferios

Paseo de la Universidad N° 300  
y Juan Díaz, Iñaquito

Quito - Ecuador.

Telf. (02) 401-4100

<https://www.uhemisferios.edu.ec/publicaciones>

Producción editorial: Diego Ortiz Jaramillo • Coordinación: Andrea Puente • Revisión Externa: Fundación Info Kratos

Esta obra fue publicada con el aval de lectores especializados, con doble par ciego, y arbitrada según las normas de publicación del Centro de Publicaciones - UHEdiciones de la Universidad Hemisferios. Queda prohibida la reproducción comercial parcial o total de esta obra, a menos que se obtenga permiso explícito del autor y la editorial.

**Cita sugerida (APA 7):**

Ullaauri Betancourt, S., Ortiz-Jaramillo, D., Fuentes Valdebenito, N., & Álvarez, F. (Coords.). (2026). *Perspectivas filosófico-políticas contemporáneas: reflexiones y debates para el siglo XXI*. UHEdiciones – Universidad Hemisferios. <https://doi.org/10.31207/uhediciones23>

# Índice

|                                                                                                                  |    |
|------------------------------------------------------------------------------------------------------------------|----|
| Prólogo .....                                                                                                    | 26 |
| Ética .....                                                                                                      | 40 |
| Presentación: la ética en la esfera pública digital: agencia,<br>poder y verdad en sociedades tecnificadas ..... | 41 |
| El espejismo del espejo: sobre qué proyectamos en las<br>máquinas .....                                          | 43 |
| La prueba del cuidado: confianza, dignidad y la tentación<br>de delegar .....                                    | 44 |
| Del cuidado al trabajo: la justicia distributiva de la<br>automatización .....                                   | 46 |
| La legitimidad epistémica: ¿por qué nadie debe custodiar<br>la verdad? .....                                     | 47 |
| ¿Quién diseña el futuro? La apertura geopolítica .....                                                           | 49 |
| El límite del medio: eficiencia, suficiencia y los tiempos del<br>antropoceno .....                              | 50 |
| Convergencias, tensiones y un programa abierto .....                                                             | 51 |



|                                                                                               |    |
|-----------------------------------------------------------------------------------------------|----|
| Sobre los límites del enfoque humanista en el desarrollo de la inteligencia artificial .....  | 55 |
| Resumen .....                                                                                 | 55 |
| Abstract .....                                                                                | 56 |
| Del autómata al algoritmo .....                                                               | 57 |
| Tendencias filosófico-políticas actuales en torno a la Inteligencia Artificial .....          | 62 |
| Crítica al enfoque humanista del desarrollo de la Inteligencia Artificial .....               | 69 |
| Argumentos a favor de la deshumanización de la tecnología .....                               | 74 |
| Conclusión .....                                                                              | 77 |
| Referencias .....                                                                             | 80 |
| Can We Trust Robots? A Brief Consideration of Human–Machine Interaction in Medical Care ..... | 84 |
| Abstract .....                                                                                | 84 |
| Resumen .....                                                                                 | 85 |
| Introduction .....                                                                            | 86 |
| Ethics of Care, with Special Reference to People with Disabilities .....                      | 89 |
| Human–Machine Interaction: Opportunities and Challenges in Healthcare .....                   | 97 |

|                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Rethinking the Concept of Trust in the Age of AI: Can We Trust Robots? .....                                                               | 106 |
| Conclusions .....                                                                                                                          | 111 |
| References .....                                                                                                                           | 112 |
| Ética de la rotación laboral en la era de la Inteligencia Artificial: ¿condición para el crecimiento económico o forma de exclusión? ..... |     |
| Resumen .....                                                                                                                              | 123 |
| Abstract .....                                                                                                                             | 124 |
| Introducción .....                                                                                                                         | 125 |
| Materiales y métodos .....                                                                                                                 | 131 |
| Desplazamiento laboral por automatización con IA ..                                                                                        | 132 |
| Beneficios laborales de la automatización con IA .....                                                                                     | 135 |
| Más allá del optimismo y pesimismo: una visión institucional .....                                                                         | 137 |
| Resultados .....                                                                                                                           | 139 |
| Conclusiones .....                                                                                                                         | 142 |
| Referencias .....                                                                                                                          | 147 |
| Razón pública y fragmentación epistémica: una defensa liberal frente a la crisis de la verdad pública .....                                |     |
| Resumen .....                                                                                                                              | 152 |
| Abstract .....                                                                                                                             | 153 |

|                                                                                                  |     |
|--------------------------------------------------------------------------------------------------|-----|
| Introducción: una variedad olvidada de crisis de legitimidad.....                                | 154 |
| La crisis epistémica como crisis de legitimidad.....                                             | 158 |
| ¿Por qué fracasa la custodia de la verdad? .....                                                 | 163 |
| La reconstrucción liberal: razón pública, confianza y virtudes.....                              | 170 |
| Objeciones, límites y conclusión .....                                                           | 180 |
| Referencias .....                                                                                | 185 |
| Transhumanismo, diseño cultural del futuro y el lugar de América Latina en el debate global..... | 190 |
| Resumen.....                                                                                     | 190 |
| Abstract .....                                                                                   | 191 |
| Introducción .....                                                                               | 192 |
| Transhumanismo.....                                                                              | 195 |
| Diseño cultural .....                                                                            | 199 |
| Mapa de actores transhumanistas .....                                                            | 202 |
| Latinoamérica frente al futuro Ttranshumanista .....                                             | 205 |
| ¿Un transhumanismo democrático y ético? Una posibilidad desde el Sur .....                       | 211 |
| Viabilidad .....                                                                                 | 218 |
| Conclusiones.....                                                                                | 226 |
| Referencias .....                                                                                | 228 |

|                                                                                                                                                                          |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| La eficiencia como categoría de la acción sostenible. Sobre la genealogía filosófica y la ética política de la indulgencia más allá del paradigma de la eficiencia ..... | 231 |
| Resumen.....                                                                                                                                                             | 232 |
| Abstract .....                                                                                                                                                           | 232 |
| Introducción .....                                                                                                                                                       | 233 |
| La engañosa promesa de la optimización en el Antropoceno .....                                                                                                           | 239 |
| Sobre la genealogía de la eficiencia. La disciplinarización de la naturaleza como proyecto filosófico de optimización .....                                              | 246 |
| Génesis filosófica moderna. La naturaleza como objeto de dominio .....                                                                                                   | 247 |
| La eficiencia como patología sociocultural en la era de la crisis climática.....                                                                                         | 251 |
| El canto de sirena de la optimización o sobre la paradoja del efecto rebote.....                                                                                         | 256 |
| La adaptación como “nueva honestidad” a la luz del rebote .....                                                                                                          | 262 |
| Conclusión y perspectivas .....                                                                                                                                          | 270 |
| Referencias .....                                                                                                                                                        | 271 |
| Pensamiento político contemporáneo .....                                                                                                                                 | 277 |

|                                                                                                        |     |
|--------------------------------------------------------------------------------------------------------|-----|
| Presentación de la sección “Pensamiento político contemporáneo” .....                                  | 278 |
| Ser, datos y poder. Una lectura crítica del sujeto datificable en la era del control algorítmico ..... | 285 |
| Resumen.....                                                                                           | 285 |
| Abstract .....                                                                                         | 286 |
| Introducción .....                                                                                     | 287 |
| Genealogía del sujeto datificable.....                                                                 | 290 |
| Del sujeto disciplinario al sujeto de rendimiento .....                                                | 290 |
| Autoexplotación, transparencia y psicopolítica .....                                                   | 295 |
| Datos, plataformas y modelación algorítmica .....                                                      | 298 |
| Smart power y conducción flexible de la conducta ....                                                  | 298 |
| Vigilancia, predicción y economía de la conducta.....                                                  | 302 |
| Ser, opacidad y crisis de la interioridad .....                                                        | 305 |
| Del dato como registro al dato como forma de mundo .....                                               | 305 |
| Libertad, juicio y autonomía bajo control algorítmico                                                  | 308 |
| Consecuencias político-democráticas .....                                                              | 312 |
| Conclusiones.....                                                                                      | 317 |
| Referencias .....                                                                                      | 321 |
| Humanismo cívico: un legado de la filosofía política clásica para el mundo contemporáneo.....          | 324 |

|                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------|-----|
| Resumen.....                                                                                                      | 324 |
| Abstract .....                                                                                                    | 325 |
| Introducción .....                                                                                                | 326 |
| Fundamentación metafísica y antropológica del<br>humanismo cívico .....                                           | 328 |
| Traducción contemporánea: instituciones, democracia y<br>esfera pública mediada .....                             | 342 |
| Concreciones éticas y políticas del humanismo cívico...                                                           | 343 |
| Objeciones actuales y respuesta del humanismo cívico.                                                             | 350 |
| El pluralismo razonable .....                                                                                     | 350 |
| La neutralidad estatal.....                                                                                       | 351 |
| El riesgo de paternalismo o perfeccionismo .....                                                                  | 352 |
| Conclusión .....                                                                                                  | 352 |
| Referencias .....                                                                                                 | 354 |
| Desigualdad neoliberal, neofascismo, tecnofeudalismo y<br>necropolítica. Claves para disputar la democracia ..... | 357 |
| Resumen.....                                                                                                      | 358 |
| Abstract .....                                                                                                    | 359 |
| Introducción .....                                                                                                | 359 |
| La desigualdad –neoliberal- es una decisión política ....                                                         | 362 |
| Neoliberalismo como ausencia de reconocimiento ....                                                               | 366 |

|                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------|-----|
| Naturalización de la desigualdad o despolitización de la<br>injusticia .....                          | 369 |
| Neofascismo y las mutaciones autoritarias del<br>neoliberalismo .....                                 | 373 |
| Tecnofeudalismo o la privación de la democracia .....                                                 | 381 |
| Neoliberalismo y necropolítica: cuando el mercado<br>decide quién vive y quién muere .....            | 385 |
| ¿Cómo disputar la democracia frente al despojo<br>neoliberal? .....                                   | 389 |
| El estado como regulador y garante del bien común .                                                   | 390 |
| Redistribuir la riqueza para reconstruir la cohesión<br>social .....                                  | 392 |
| Acceso a conocimiento y educación como pilares<br>centrales de la democracia .....                    | 395 |
| Repolitizar la economía para desmontar el régimen<br>ideológico neoliberal .....                      | 397 |
| Desmercantilizar la vida social, reconstruir lo común                                                 | 399 |
| Hacia una democracia postneoliberal basada en la<br>justicia, la cooperación y la sostenibilidad..... | 401 |
| Conclusión .....                                                                                      | 404 |
| Referencias .....                                                                                     | 407 |
| Videos: .....                                                                                         | 410 |
| Prensa: .....                                                                                         | 410 |

|                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------|-----|
| Inteligencia artificial y tecnoceno: genealogía, control y dilemas éticos de la revolución algorítmica ..... | 413 |
| Resumen.....                                                                                                 | 413 |
| Abstract .....                                                                                               | 414 |
| Introducción .....                                                                                           | 415 |
| La revolución algorítmica .....                                                                              | 420 |
| La cultura del algoritmo.....                                                                                | 423 |
| Por una ética de la Inteligencia Artificial .....                                                            | 430 |
| El aceleracionismo: hacia una teoría de la fluidez radical de la IA .....                                    | 440 |
| Conclusión .....                                                                                             | 448 |
| Referencias .....                                                                                            | 449 |
| Gubernamentalidad algorítmica: líneas de fuga para una crítica del poder digital contemporáneo .....         | 456 |
| Resumen.....                                                                                                 | 457 |
| Abstract .....                                                                                               | 458 |
| Introducción .....                                                                                           | 459 |
| Racionalidad algorítmica: del perfil al gobierno digital. ....                                               | 461 |
| Líneas de fuga y expresiones actuales para entender la GA .....                                              | 465 |
| Primera línea estructurante: subjetivación, biopolítica y delegación de decisiones .....                     | 466 |



|                                                                                                 |     |
|-------------------------------------------------------------------------------------------------|-----|
| Segunda línea estructurante: vigilancia y control<br>algorítmico .....                          | 471 |
| Tercera línea estructurante: opacidad, transparencia y<br>manipulación .....                    | 476 |
| Cuarta línea estructurante: desigualdad, exclusión y<br>resistencias.....                       | 481 |
| Articulaciones estructurales: la economía de plataformas<br>como epicentro tensional.....       | 486 |
| Consideraciones finales: gubernamentalidad algorítmica<br>en acto y horizontes por pensar ..... | 489 |
| Referencias .....                                                                               | 492 |
| Mecanismos de participación ciudadana en la sociedad<br>digital .....                           | 503 |
| Resumen.....                                                                                    | 503 |
| Abstract .....                                                                                  | 504 |
| Introducción .....                                                                              | 505 |
| Relevancia de la participación ciudadana en la era<br>digital .....                             | 507 |
| Definición de participación ciudadana y sociedad digital<br>.....                               | 509 |
| Relación entre democracia participativa, gobernanza<br>digital y ciudadanía activa .....        | 512 |

|                                                                                                                                                                  |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Mecanismos de participación ciudadana en la sociedad digital.....                                                                                                | 514 |
| Plataformas de gobierno electrónico .....                                                                                                                        | 514 |
| Acceso a la información y transparencia digital.....                                                                                                             | 517 |
| Retos y desafíos .....                                                                                                                                           | 519 |
| Brecha digital y exclusión en la participación política                                                                                                          | 522 |
| Derechos digitales y protección de datos .....                                                                                                                   | 524 |
| Conclusiones y perspectivas.....                                                                                                                                 | 527 |
| Referencias .....                                                                                                                                                | 529 |
| Emociones y movilización política en el sexenio de AMLO (2018-2024). La reconfiguración de la política mexicana en el contexto de la Cuarta Transformación ..... | 537 |
| Resumen.....                                                                                                                                                     | 537 |
| Abstract .....                                                                                                                                                   | 538 |
| Introducción .....                                                                                                                                               | 539 |
| La política y la emoción en la sociedad contemporánea                                                                                                            | 544 |
| La subjetividad en el contexto de la cuarta transformación .....                                                                                                 | 553 |
| Conclusión .....                                                                                                                                                 | 566 |
| Referencias .....                                                                                                                                                | 569 |
| Itinerarios del pensamiento político y las hegemonías en la configuración del sujeto social.....                                                                 | 577 |

|                                                                                                                                                   |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Resumen.....                                                                                                                                      | 577 |
| Abstract .....                                                                                                                                    | 578 |
| Introducción .....                                                                                                                                | 579 |
| Fundamentación metodológica.....                                                                                                                  | 581 |
| La polis griega como matriz originaria.....                                                                                                       | 583 |
| El tránsito hacia el Estado moderno .....                                                                                                         | 587 |
| Ideologías democráticas renovadas .....                                                                                                           | 591 |
| Género, comunidad y emancipación del sujeto social ...                                                                                            | 595 |
| Epistemología social y ética en el itinerario político y la<br>difusión del conocimiento.....                                                     | 599 |
| Continuidades y rupturas en la historia del pensamiento<br>político.....                                                                          | 603 |
| Conclusiones.....                                                                                                                                 | 607 |
| Referencias .....                                                                                                                                 | 609 |
| Cristianismo y democracia moderna: la influencia de la<br>doctrina social de la Iglesia en la construcción de<br>instituciones democráticas ..... | 619 |
| Resumen.....                                                                                                                                      | 619 |
| Abstract .....                                                                                                                                    | 620 |
| Introducción .....                                                                                                                                | 621 |
| Cristianismo y democracia moderna.....                                                                                                            | 625 |
| La doctrina cristiana en la esfera pública .....                                                                                                  | 633 |

|                                                                                                                       |     |
|-----------------------------------------------------------------------------------------------------------------------|-----|
| La ciudadanía cristiana .....                                                                                         | 640 |
| Conclusión .....                                                                                                      | 643 |
| Referencias .....                                                                                                     | 646 |
| La crítica rapsódica de Rawls al capitalismo: un fundamento<br>antropológico para la democracia de propietarios ..... | 651 |
| Resumen .....                                                                                                         | 652 |
| Abstract .....                                                                                                        | 653 |
| Introducción .....                                                                                                    | 653 |
| La crítica rapsódica de Rawls al capitalismo: fragmentos<br>de un diagnóstico estructural .....                       | 657 |
| El constructivismo kantiano como fundamento<br>antropológico .....                                                    | 661 |
| Los poderes morales como estructura antropológica<br>fundamental .....                                                | 662 |
| La distinción racional/razonable como estructura de la<br>agencia moral .....                                         | 664 |
| El constructivismo como alternativa a<br>fundamentaciones epistemológica .....                                        | 666 |
| La autonomía como principio unificador .....                                                                          | 670 |
| Pluralismo como consecuencia de la formalidad<br>autónoma .....                                                       | 672 |
| La posición original como dispositivo de representación<br>.....                                                      | 674 |

|                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Los tres niveles de publicidad y la objetividad<br>construida a través del equilibrio reflexivo .....                                      | 674 |
| El equilibrio reflexivo como realización metodológica de<br>la antropología rawlsiana .....                                                | 676 |
| Conclusión: la antropología rawlsiana como fundamento<br>de la democracia de propietarios a través del equilibrio<br>reflexivo .....       | 679 |
| La ambigüedad en la propiedad y su resolución<br>institucional .....                                                                       | 682 |
| La superioridad de la democracia de propietarios sobre<br>el socialismo de mercado obligatorio desde la<br>perspectiva antropológica ..... | 685 |
| La democracia de propietarios como realización de la<br>reciprocidad y los intereses de orden superior .....                               | 688 |
| El principio de reciprocidad como mandato constitutivo<br>de la estructura antropológica .....                                             | 691 |
| Respuesta a la crítica de G. A. Cohen: las instituciones<br>como realización del ethos derivado de la antropología<br>.....                | 692 |
| Ideas finales .....                                                                                                                        | 695 |
| Referencias .....                                                                                                                          | 697 |
| Gobernanza ética y psicopolítica digital: razón algorítmica,<br>poder y circulación social del conocimiento .....                          | 701 |
| Resumen .....                                                                                                                              | 701 |

|                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------|-----|
| Abstract .....                                                                                                | 702 |
| Introducción .....                                                                                            | 703 |
| Marco teórico.....                                                                                            | 706 |
| Definiciones operativas de los conceptos fundamentales<br>.....                                               | 706 |
| Una crítica filosófica a la digitalización .....                                                              | 709 |
| Epistemología social de la razón algorítmica.....                                                             | 712 |
| Planteamiento del problema y pregunta de investigación<br>.....                                               | 719 |
| Metodología.....                                                                                              | 721 |
| Resultados hermenéuticos .....                                                                                | 726 |
| El diagnóstico de la sociedad digital a través de la IA                                                       | 726 |
| La tiranía de la visibilidad: transparencia algorítmica y<br>control psicopolítico.....                       | 727 |
| La erosión del ser: hiperconexión, fatiga informativa y el<br>eclipse de la razón comunicativa .....          | 729 |
| La dictadura de la eficiencia: positividad,<br>autoexplotación y la automatización de la desigualdad<br>..... | 733 |
| Discusión e implicaciones .....                                                                               | 736 |
| Hacia una gobernanza ética y una resistencia humana:<br>un marco de respuesta multinivel .....                | 736 |

|                                                                                                                                                                         |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Limitaciones y futuras líneas de investigación .....                                                                                                                    | 741 |
| Conclusiones.....                                                                                                                                                       | 743 |
| Referencias .....                                                                                                                                                       | 748 |
| Trade, Security, and Selective Alignment: Process Tracing<br>the Peru-United States Trade Promotion Agreement and the<br>Reorientation of Peruvian Foreign Policy ..... | 752 |
| Abstract .....                                                                                                                                                          | 753 |
| Resumen.....                                                                                                                                                            | 754 |
| Introducción .....                                                                                                                                                      | 755 |
| Literature Review and Theoretical Framework.....                                                                                                                        | 759 |
| Research Design .....                                                                                                                                                   | 765 |
| From ATPA to the PTPA .....                                                                                                                                             | 771 |
| Process-Tracing the Agreement .....                                                                                                                                     | 777 |
| Mechanism 1 - Preference Insecurity and Coalition<br>Formation in Peru .....                                                                                            | 777 |
| Mechanism 2 - U.S. Competitive Liberalization and<br>Andean Strategy .....                                                                                              | 781 |
| Mechanism 3 - Deep Agreement and Institutional Lock-<br>In.....                                                                                                         | 783 |
| Foreign-Policy Effects of the Agreement.....                                                                                                                            | 786 |
| Counternarcotics and the Security Agenda.....                                                                                                                           | 786 |

|                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------|-----|
| Regional Integration and the Fragmentation of the Andean Space.....                                            | 787 |
| Peru and the Bolivarian Challenge .....                                                                        | 789 |
| Economic Outcomes, Distribution, and Structural Limits .....                                                   | 791 |
| Discussion.....                                                                                                | 797 |
| Conclusion .....                                                                                               | 802 |
| References .....                                                                                               | 804 |
| Epistemología Social .....                                                                                     | 811 |
| Presentación de la sección “Epistemología social” .....                                                        | 812 |
| Responsabilidades éticas en la era de la posverdad: desafíos epistémicos en las redes sociales digitales ..... | 817 |
| Resumen.....                                                                                                   | 817 |
| Abstract .....                                                                                                 | 818 |
| Introducción .....                                                                                             | 819 |
| Fundamentos éticos en la difusión del conocimiento digital.....                                                | 823 |
| Desafíos epistémicos en la posverdad .....                                                                     | 828 |
| Propuestas para una comunicación responsable.....                                                              | 833 |
| Consideraciones para contextos latinoamericanos ....                                                           | 837 |
| Conclusiones.....                                                                                              | 841 |
| Referencias .....                                                                                              | 844 |



|                                                                                                                      |     |
|----------------------------------------------------------------------------------------------------------------------|-----|
| El conocimiento como bien común: una reflexión ética sobre su difusión en la era de la inteligencia artificial ..... | 849 |
| Resumen .....                                                                                                        | 849 |
| Abstract .....                                                                                                       | 850 |
| Introducción .....                                                                                                   | 851 |
| La responsabilidad epistémica como concepto clave.....                                                               | 856 |
| Entre Fricker y Medina: injusticia epistémica estructural .....                                                      | 859 |
| Goldman, Latour y la fiabilidad distribuida.....                                                                     | 860 |
| Inteligencia Artificial y nuevas formas de acceso al conocimiento .....                                              | 861 |
| Plataformas basadas en IA y circulación del saber ....                                                               | 863 |
| Oportunidades y riesgos: el rol de los algoritmos en la mediación cognitiva .....                                    | 864 |
| El conocimiento entre el bien común y la mercantilización .....                                                      | 867 |
| Modelos de negocio y barreras de acceso .....                                                                        | 868 |
| Acceso abierto como principio ético .....                                                                            | 870 |
| Repositorios abiertos, saber compartido y equidad epistémica.....                                                    | 871 |
| Plataformas como arXiv, CORE, PubMed Central, DOAJ .....                                                             | 873 |

|                                                                                                           |     |
|-----------------------------------------------------------------------------------------------------------|-----|
| El conocimiento como patrimonio colectivo.....                                                            | 874 |
| Conclusiones y proyecciones éticas .....                                                                  | 876 |
| Propuesta de principios para una ética digital del<br>conocimiento.....                                   | 879 |
| El papel ambivalente de la IA: ¿aliada o amenaza? ...                                                     | 880 |
| Referencias .....                                                                                         | 882 |
| Ética y conocimiento: el problema de la discriminación de<br>los testimonios y los <i>deepfakes</i> ..... | 886 |
| Resumen.....                                                                                              | 886 |
| Abstract .....                                                                                            | 887 |
| Introducción .....                                                                                        | 888 |
| ¿El testimonio como fuente de conocimiento? .....                                                         | 889 |
| Discriminar la palabra de otros.....                                                                      | 893 |
| La credibilidad y nuevas tecnologías: <i>deepfakes</i> .....                                              | 898 |
| Conclusiones.....                                                                                         | 903 |
| Referencias .....                                                                                         | 903 |
| <i>Deepfake</i> y regímenes de visibilidad: Lula, Bolsonaro y el<br>gesto inexistente.....                | 907 |
| Resumen.....                                                                                              | 907 |
| Abstract .....                                                                                            | 908 |
| Introducción .....                                                                                        | 909 |

|                                                                                                                                                                              |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Fundamentación teórica .....                                                                                                                                                 | 915 |
| Imagen de una reconciliación fabulada.....                                                                                                                                   | 920 |
| La audiencia y las disputas de interpretación .....                                                                                                                          | 926 |
| Consideraciones finales .....                                                                                                                                                | 929 |
| Referencias .....                                                                                                                                                            | 931 |
| “I don't like them putting fraud in my conspiracy theories!”:<br>A Philosophical Examination of Conspiracy Theories,<br>'Pataphysics, and Alex Jones's Frog Conspiracy ..... | 937 |
| Abstract .....                                                                                                                                                               | 939 |
| Resumen.....                                                                                                                                                                 | 940 |
| Introducción .....                                                                                                                                                           | 941 |
| The Concepts of Conspiracy and 'Pataphysics.....                                                                                                                             | 943 |
| What is a Conspiracy Theory? .....                                                                                                                                           | 943 |
| What is a Pataphysical Theory? .....                                                                                                                                         | 948 |
| Alex Jones's Frog Conspiracy, a Case Study .....                                                                                                                             | 953 |
| Conclusions .....                                                                                                                                                            | 960 |
| References .....                                                                                                                                                             | 963 |
| El auge de las revistas de baja calidad en el ámbito<br>académico: causas, consecuencias y desafíos éticos.....                                                              | 968 |
| Resumen.....                                                                                                                                                                 | 968 |
| Abstract .....                                                                                                                                                               | 970 |

|                                                                               |      |
|-------------------------------------------------------------------------------|------|
| Introducción .....                                                            | 971  |
| Método .....                                                                  | 978  |
| El ecosistema de las revistas científicas globalizadas ....                   | 979  |
| La indexación en bases de datos y su rol en el<br>ecosistema científico ..... | 982  |
| Las revistas científicas legítimas .....                                      | 984  |
| El proceso de publicación en las revistas científicas ..                      | 991  |
| <i>Las revistas depredadoras: tipos, causas y consecuencias</i><br>.....      | 997  |
| Revistas fraudulentas o de estafa. ....                                       | 997  |
| Conclusiones.....                                                             | 1006 |
| Causas y consecuencias del auge de las revistas de baja<br>calidad.....       | 1006 |
| Importancia del control de calidad en las bases de datos<br>académicas .....  | 1007 |
| Relevancia del ORCID en la credibilidad académica                             | 1008 |
| Publicar: un proceso de aprendizaje y consolidación<br>académica .....        | 1008 |
| Construcción de un perfil académico y oportunidades<br>como revisor.....      | 1009 |
| Publicar como un proceso de desarrollo.....                                   | 1010 |

|                                                                      |      |
|----------------------------------------------------------------------|------|
| Desafío colectivo: hacia una ciencia ética y comprometida .....      | 1011 |
| Referencias .....                                                    | 1011 |
| Understanding as a Requisite for Institutional Trustworthiness ..... | 1017 |
| Abstract .....                                                       | 1018 |
| Resumen .....                                                        | 1019 |
| Introduction .....                                                   | 1019 |
| Robustness .....                                                     | 1021 |
| Self-advertisement .....                                             | 1023 |
| Transparency .....                                                   | 1025 |
| Understanding and Trust .....                                        | 1032 |
| Institutions and Understanding .....                                 | 1037 |
| Institutional Understanding .....                                    | 1038 |
| How Institutions Manifest Understanding .....                        | 1042 |
| How Institutions Make Themselves Understandable .....                | 1046 |
| Social problems and value alignment .....                            | 1050 |
| Conclusion .....                                                     | 1052 |
| References .....                                                     | 1053 |



# Prólogo

## Pensar en medio de la intemperie

Hay libros que nacen como una acumulación de respuestas y hay otros, acaso los más perentorios, que surgen de la imposibilidad de seguir eludiendo ciertas incógnitas. *Perspectivas filosófico-políticas contemporáneas: reflexiones y debates para el siglo XXI* pertenece, indubitadamente, a esta segunda estirpe. No comparece ante el lector como un repertorio aséptico de disquisiciones académicas, ni como una colección de textos yuxtapuestos por la mera vecindad temática. Se presenta, más bien, como una cartografía de nuestra perplejidad, como un intento plural, abigarrado y rigurosamente deliberativo por comprender una época en la que las viejas categorías conservan su nombre mientras sus contenidos se metamorfosean ante nuestros ojos.

Este volumen encarna, además, una de las virtudes más fecundas de la vida universitaria, aquella que consiste en reunir voces divergentes sin obligarlas a disolverse en una

unanimidad impostada. La universidad cumple su función más elevada cuando, en lugar de administrar certezas prefabricadas, habilita espacios donde las ideas pueden encontrarse, contradecirse y depurarse en el crisol de la argumentación. Este libro es resultado de esa vocación, y su mérito, lejos de residir en clausurar los debates que aborda, estriba en devolverles la densidad que la premura digital, la polarización política y la banalización discursiva suelen arrebatarnos.

Vivimos bajo el signo de una paradoja monumental. Nunca habíamos dispuesto de tantos medios para producir, almacenar y transmitir conocimiento y, sin embargo, pocas veces pareció tan lábil nuestra capacidad para distinguir lo verdadero de lo verosímil, la evidencia de lo estentóreo o el argumento de la mera reiteración algorítmica. La información se reproduce con una exuberancia casi oceánica, pero la comprensión continúa siendo un bien escaso. Las sociedades están hiperconectadas sin ser por ello más comunicadas, los sujetos se hallan permanentemente expuestos a la palabra ajena y cada vez menos dispuestos a escucharla y comprenderla, y las instituciones proclaman una transparencia meridiana mientras las arquitecturas tecnológicas que ordenan la visibilidad permanecen envueltas en una secrecía casi sacerdotal.

En medio de este escenario, la pregunta por la difusión del conocimiento deja de ser un asunto exclusivamente

epistemológico y se convierte también en una pregunta ética, porque concierne a nuestra responsabilidad ante los demás, en una pregunta política, porque involucra la distribución del poder, y en una pregunta antropológica, porque termina interrogando aquello que todavía entendemos por humanidad. ¿Quién puede hablar? ¿A quién se concede credibilidad? ¿Qué voces son amplificadas y cuáles condenadas a la penumbra? ¿Quién diseña los sistemas que jerarquizan el saber? ¿Bajo qué criterios una máquina recomienda, clasifica, descarta o predice? Y, sobre todo, ¿qué ocurre con la libertad cuando las condiciones de nuestra elección han sido configuradas de antemano por mecanismos que apenas comprendemos?

El eje ético del volumen se adentra con particular agudeza en estas interrogantes. Sus páginas obligan a abandonar la epidérmica dicotomía entre tecnofilia y tecnofobia, entre la celebración irreflexiva de la innovación y el apocalipticismo que atribuye a cada máquina una inminente vocación tiránica. La técnica, lejos de aparecer aquí como un ente metafísico dotado de voluntad propia, se revela como una construcción humana atravesada por decisiones, intereses, valores y omisiones.

Una primera provocación consiste en cuestionar la obsesión antropomórfica que ha acompañado el desarrollo de la inteligencia artificial. Durante siglos hemos imaginado la máquina como espejo, facsímil o sucedáneo del ser humano,



y hemos aspirado a verla pensar, hablar, crear y comportarse como nosotros, como si la única inteligencia concebible fuese aquella que reproduce nuestras formas cognitivas. Frente a ello, el libro invita a preservar la alteridad de lo tecnológico y a reconocer que una máquina no necesita convertirse en un remedo de humanidad para ser útil, pues, paradójicamente, cuanto más confundimos sus operaciones con atributos humanos, mayor es el riesgo de abdicar de nuestra responsabilidad moral. Deshumanizar la tecnología no significa despojarla de límites éticos, pues equivale más bien a dejar de proyectar sobre ella cualidades que pertenecen al juicio, a la empatía, a la vulnerabilidad y a la experiencia encarnada.

Esa frontera adquiere una gravedad inmediata cuando la inteligencia artificial y la robótica ingresan en el ámbito del cuidado médico. La posibilidad de que los robots acompañen a personas mayores, a pacientes con demencia o a individuos en condiciones de dependencia no puede evaluarse exclusivamente desde la eficiencia, porque el cuidado nunca fue una concatenación de tareas mecánicas, en la medida en que involucra confianza, reconocimiento, cariño, reciprocidad y presencia. La máquina puede asistir, vigilar guarismos, recordar medicamentos o aliviar parte de la carga física, pero la delegación irrestricta amenaza con convertir una crisis social del cuidado en una solución tecnológica estratificada, donde unos reciben proximidad humana y otros apenas su simulacro automatizado. El problema decisivo,

entonces, más que determinar si un robot puede ejecutar una función, exige preguntarnos qué clase de sociedad estamos construyendo cuando decidimos confiarle nuestra vulnerabilidad.

La misma ambivalencia se proyecta sobre el trabajo. La inteligencia artificial puede incrementar la productividad, estimular nuevas ocupaciones y liberar a las personas de tareas repetitivas, aunque también puede acelerar desplazamientos laborales, profundizar desigualdades y condenar a ciertos sectores a una obsolescencia decretada desde los centros de innovación. El crecimiento económico, desprovisto de instituciones inclusivas, puede convertirse en un eufemismo para nombrar la exclusión, y la rotación laboral dista de ser éticamente neutral cuando sus costos recaen sobre quienes poseen menores posibilidades de reconversión. Allí donde la técnica multiplica las capacidades productivas, la política debe preguntarse quién recibe los beneficios, quién soporta las pérdidas y qué mecanismos de justicia distributiva impiden que la innovación sea sufragada mediante la precarización de los más vulnerables.

El debate alcanza después el corazón mismo de la democracia, que reside en su legitimidad epistémica. Una comunidad política no se sostiene únicamente mediante elecciones, leyes y procedimientos, pues necesita también un suelo compartido de confianza que permita reconocer razones, confrontar evidencias y corregir errores. Cuando ese

suelo se resquebraja, la falsedad deja de ser una anomalía ocasional y se convierte en un método de fragmentación. La respuesta, con todo, no puede consistir en instaurar un custodio onnipotente de la verdad, ya que ni el Estado, ni una tecnocracia, ni las grandes plataformas deben subsumir el juicio ciudadano bajo la pretensión de protegerlo. La democracia debe cuidar la arquitectura del espacio público, su pluralismo, su transparencia y su apertura, sin erigirse en soberana omnímoda de sus contenidos, porque proteger las condiciones del juicio no equivale a suplantarlo.

Finalmente, el horizonte ético se expande hacia la disputa geopolítica por el futuro. Las promesas transhumanistas, los proyectos de mejoramiento humano y las narrativas de superación de la fragilidad suelen presentarse como aspiraciones universales, aunque muchas veces hayan sido concebidas desde ecosistemas tecnocientíficos extraordinariamente localizados. América Latina no puede limitarse a contemplar este porvenir como consumidora periférica y debe intervenir en su definición, introducir la memoria de la colonialidad y reivindicar la sostenibilidad, el Buen Vivir, la justicia estructural y la pluralidad de cosmologías. El futuro tecnológico no pertenece únicamente a quienes pueden financiarlo, pues pertenece también a quienes deberán habitar sus consecuencias.

El segundo gran itinerario del libro recorre las convulsiones del pensamiento político contemporáneo, y lo

hace sin incurrir en la arrogancia presentista de suponer que todo cuanto nos ocurre carece de precedentes. La polis griega, la virtud aristotélica, la amistad cívica y el bien común reaparecen, lejos de toda condición de reliquia ornamental, como recursos conceptuales para examinar la anemia comunitaria de nuestro tiempo. El humanismo cívico recuerda que la libertad no se agota en la ausencia de interferencia y que la ciudadanía exige formación del carácter, disposición hacia lo común y capacidad para deliberar sobre los fines colectivos.

En esa reconstrucción histórica, el cristianismo y la doctrina social de la Iglesia son examinados en su compleja relación con la democracia moderna. Lejos de una narración simplista de antagonismo o de perfecta continuidad, se estudia el tránsito desde la inicial desconfianza hacia la modernidad hasta la contribución de categorías como dignidad humana, solidaridad, subsidiariedad y bien común. La democracia aparece así no solo como un procedimiento para contabilizar voluntades, también como un orden político que requiere un horizonte moral capaz de impedir que la mayoría se convierta en mera aritmética del poder.

También Rawls regresa a escena. Su crítica al capitalismo de bienestar, aparentemente dispersa o rapsódica, encuentra una posible coherencia en la exigencia de que la libertad política posea condiciones materiales efectivas. La democracia de propietarios se plantea como una

alternativa que, antes que limitarse a corregir desigualdades después de que el mercado las produzca, aspira a predistribuir recursos y capacidades para evitar la concentración que vuelve ilusoria la igualdad ciudadana. La autonomía, bajo esta lectura, no habita en un vacío institucional, pues necesita bases sociales y económicas que permitan a las personas ejercer sus poderes morales sin quedar sometidas a dependencias arbitrarias.

Junto a estas genealogías, el libro despliega una crítica con acritud de las mutaciones contemporáneas del poder. La desigualdad neoliberal, el neofascismo, la necropolítica y el tecnofeudalismo son estudiados como expresiones de un orden que concentra riqueza, privatiza capacidades colectivas y administra diferencialmente la precariedad. Recuperar el Estado, en este contexto, no significa idealizarlo ni ignorar su historia de exclusiones y los defectos que de él dimanan, pues supone disputar su orientación y reconstruirlo como garante de derechos, de redistribución y de protección democrática.

Pero el poder no actúa únicamente mediante instituciones, coerción o recursos materiales, ya que también gobierna afectos. El análisis de la movilización política durante el sexenio de Andrés Manuel López Obrador muestra cómo la esperanza, el resentimiento, la redención moral, el amor al pueblo y la construcción afectiva del adversario pueden reorganizar identidades políticas. La emocionalización no constituye una anomalía exterior a la

política, pues forma parte de su sustancia, y el riesgo aparece cuando el antagonismo moral reemplaza la discusión programática y cuando el ciudadano deja de ser un interlocutor para convertirse en un sujeto permanentemente movilizado por adhesiones y repulsiones.

Asimismo, el estudio del acuerdo comercial entre Perú y Estados Unidos recuerda que la política exterior y la economía internacional nunca son dominios puramente técnicos. Un tratado comercial puede funcionar como mecanismo de alineamiento estratégico, consolidar modelos económicos y elevar los costos de futuras transformaciones regulatorias. Los Estados periféricos conservan agencia, aunque la ejercen dentro de estructuras profundamente asimétricas, y esa tensión entre decisión y condicionamiento atraviesa buena parte de la experiencia latinoamericana.

La sociedad digital complejiza todavía más estos dilemas. Las tecnologías amplían los canales de participación ciudadana, facilitan la organización y aproximan a los individuos a las instituciones, al tiempo que reproducen brechas, promueven formas superficiales de intervención y entregan a actores privados la infraestructura de la esfera pública. La participación digital solo fortalece la democracia cuando se acompaña de acceso equitativo, alfabetización tecnológica, transparencia y garantías para los derechos fundamentales.

Aquí convergen las reflexiones sobre el tecnoceno, la psicopolítica y la gubernamentalidad algorítmica. El poder digital ya no necesita ordenar de manera explícita cuando puede predecir, sugerir, personalizar y modular, pues gobierna no solamente los cuerpos, también la atención, los deseos y las probabilidades de conducta, y convierte la identidad en perfil, la experiencia en dato y la deliberación en una sucesión de estímulos optimizados.

En este punto se inscribe una de las meditaciones más hondas del volumen, la que interroga la relación entre el ser y los datos. La datificación no constituye una operación meramente técnica de registro, dado que encierra una auténtica mutación ontológica y política, pues aquello que la tradición filosófica pensó como persona, interioridad o existencia encarnada comparece ahora como un flujo de información cuantificable, agregable y explotable. El sujeto datificable ya no es interpelado como un interlocutor que responde de sí, puesto que resulta descompuesto en huellas infraindividuales y recompuesto en perfiles supraindividuales que anticipan su conducta sin requerir su palabra ni su consentimiento reflexivo. Cuando el ser queda reducido a aquello que de él puede medirse, la libertad cede su lugar a la probabilidad y la biografía se disuelve en la correlación estadística. Frente a esa reducción, estas páginas reivindican la opacidad constitutiva de la persona, aquella reserva de interioridad que ningún registro agota y que permanece como condición de toda subjetivación política genuina. El libro, con

todo, no abandona al sujeto dentro de la maquinaria que lo computa, ya que busca líneas de fuga, espacios de resistencia, auditorías, mecanismos de rendición de cuentas y formas de gobernanza ética capaces de devolver la tecnología al terreno de la decisión democrática.

El tercer eje, dedicado a la epistemología social, completa y a la vez desborda los anteriores. La posverdad es estudiada, antes que como una simple proliferación de mentiras, como la erosión de las responsabilidades que sostienen el conocimiento colectivo. En las redes digitales, usuarios, plataformas, medios e instituciones participan de una ecología en la que cada amplificación, silenciamiento y jerarquización posee consecuencias morales, y la verdad pública no puede sobrevivir cuando todos reclaman el derecho a hablar pero nadie acepta el deber de responder por lo dicho.

El conocimiento, bajo esta perspectiva, debe ser defendido como bien común. La inteligencia artificial puede democratizar el acceso, derribar barreras lingüísticas y facilitar la comprensión, aunque también puede homogeneizar el pensamiento, invisibilizar saberes y concentrar el poder epistémico. La transparencia algorítmica, la pluralidad cognitiva, el acceso abierto y la responsabilidad compartida dejan de ser aspiraciones accesorias y se convierten en condiciones de una ciudadanía capaz de conocer y participar.



Los capítulos dedicados a los deepfakes muestran la profundidad de esta crisis. La imagen sintética no solo falsifica acontecimientos, puede además degradar el testimonio, perjudicar la credibilidad de una persona y reordenar los afectos públicos alrededor de escenas que jamás ocurrieron. El abrazo inexistente entre Lula y Bolsonaro revela que lo falso puede producir efectos políticos verdaderos y activar deseos de reconciliación, repulsiones, burlas o esperanzas, de modo que la frontera entre realidad y posibilidad se torna porosa y con ella se tambalea la confianza en la imagen como evidencia.

Las teorías conspirativas comparecen, asimismo, despojadas de su aparente inocencia excéntrica, ya que, bajo ciertas condiciones, la extravagancia puede servir como vehículo para el fraude, la manipulación y el lucro a costa de la credulidad. Algo semejante ocurre dentro de la propia academia cuando la presión por publicar, la mercantilización del conocimiento y la proliferación de revistas de baja calidad erosionan los mecanismos que deberían garantizar la integridad científica. La crisis de la verdad no habita únicamente en los márgenes de internet, pues puede instalarse también en las instituciones encargadas de producir conocimiento legítimo.

Por ello, el recorrido concluye con una pregunta fundamental acerca de qué hace confiable a una institución. No basta con que se proclame experta, transparente o

virtuosa, dado que debe comprender los problemas que enfrenta, manifestar ese entendimiento y hacerse comprensible para quienes dependen de ella, porque no puede existir confianza genuina en aquello que se presenta como una caja negra, trátase de una universidad, de un organismo estatal, de una revista científica o de un algoritmo.

Así, los tres ejes de este libro terminan confluyendo en una misma tríada de dignidad, agencia y verdad. La dignidad impide reducir a la persona a dato, a consumidor, a paciente pasivo o a trabajador prescindible, mientras que la agencia exige que nadie sea tratado únicamente como objeto de administración, persuasión o predicción, y la verdad proporciona el mundo común sin el cual la deliberación se disuelve en una guerra de simulacros.

El lector no encontrará aquí una doctrina monolítica. Hallará tensiones entre liberalismo y comunitarismo, entre universalidad y pluralidad, entre innovación y prudencia, entre autonomía y bien común, entre objetividad y experiencia situada, y esa heterogeneidad, lejos de constituir una debilidad editorial, es la expresión honesta de un tiempo cuyas contradicciones no pueden resolverse mediante una sola tradición teórica.

En una época seducida por la velocidad, este libro reivindica la pausa del pensamiento. Frente a la grandilocuencia tecnológica recupera la pregunta ética, frente

a la crispación política restaura la complejidad y frente a la vorágine informativa exige comprensión. Su invitación última no consiste en temer al futuro ni en rendirse ante él, pues nos convoca a disputarlo, porque la tecnología, la democracia y el conocimiento, lejos de ser destinos inexorables, son construcciones humanas y, por tanto, permanecen abiertas a nuestra responsabilidad.

Que estas páginas sirvan entonces como umbral, no como pórtico ornamental, como invitación a ingresar en una conversación incómoda y necesaria. Pensar el siglo XXI implica reconocer que habitamos una intemperie conceptual, política y moral, aunque implica también recordar que ninguna intemperie es absoluta mientras subsista la voluntad de encontrarnos en la palabra, de justificar nuestras decisiones ante los demás y de construir un mundo común que no sacrifique la dignidad en el altar de la eficiencia.

*Roy Alexander Torres Anangonó*

*Facultad Latinoamericana de Ciencias Sociales - Ecuador*



# Ética

# **Presentación: la ética en la esfera pública digital: agencia, poder y verdad en sociedades tecnificadas**

**Nicolás Fuentes Valdebenito<sup>1</sup>**

Hay una imagen, antigua y persistente, según la cual la ética es ante todo un asunto de la proximidad: el rostro del otro, la promesa empeñada, el daño que una persona inflige a otra dentro del estrecho radio de su alcance corporal. Esa imagen no es falsa, pero ha dejado de ser suficiente. Las relaciones morales que organizan hoy la vida en común no transcurren solo entre individuos que se miran a los ojos, sino a través de plataformas que median la palabra pública, de sistemas algorítmicos que distribuyen la visibilidad y la

---

<sup>1</sup> Editor responsable del eje Ética. Director Ejecutivo de la Fundación CIEPE — Centro Interdisciplinario en Ética, Política y Economía. <https://orcid.org/0000-0002-2229-7488>

sospecha, de arquitecturas informacionales que configuran de antemano qué puede ser dicho, creído y compartido, y de infraestructuras materiales cuya lógica de funcionamiento condiciona, en última instancia, qué futuros son siquiera sostenibles. La agencia se ejerce dentro de esas estructuras; el poder se concentra en quienes las diseñan; la vulnerabilidad se redistribuye según lógicas que ningún agente individual controla del todo. Reunir bajo el rótulo de ética los cinco capítulos que componen este eje supone, por ello, una tesis previa: que la reflexión moral contemporánea no puede pensarse ya únicamente desde el vínculo interpersonal, sino que debe asumir como objeto propio las mediaciones técnicas, y los límites materiales que estas mediaciones presuponen, que constituyen el espacio donde ese vínculo se vuelve posible o imposible.

Los trabajos aquí reunidos no comparten una escuela ni una conclusión. Proviene de la filosofía de la tecnología, de la bioética del cuidado, de la economía política del trabajo, de la teoría liberal de la legitimidad, de la crítica decolonial del futuro y de la filosofía política de la sostenibilidad. Sin embargo, leídos en conjunto, dibujan un mismo problema desde ángulos distintos: el de la redistribución de la agencia moral en un mundo donde lo humano coexiste, colabora y compete con artefactos que actúan, y donde esa coexistencia encuentra además un límite que ningún diseño institucional puede eludir: el de los recursos finitos sobre los que toda acción, humana o artefactual, se sostiene. La pregunta que

recorre el eje no es si la técnica es buena o mala (dicotomía que todos los autores rechazan), sino bajo qué condiciones la mediación tecnológica amplía o erosiona aquello que hace de un ser humano un sujeto: su capacidad de juzgar, de confiar fundadamente, de ser reconocido como fuente de razones y no como objeto de gestión, y bajo qué condiciones materiales esa misma capacidad puede ejercerse en el tiempo, sin hipotecar las posibilidades de quienes vendrán después.

### **El espejismo del espejo: sobre qué proyectamos en las máquinas**

El primer movimiento del eje es de orden conceptual, y consiste en desactivar una confusión fundacional. Pablo de Miguel, en su examen de los límites del enfoque humanista en el desarrollo de la inteligencia artificial, traza una genealogía que va de los autómatas de Hefesto al *hombre máquina* de La Mettrie y de allí a los sistemas algorítmicos contemporáneos, para mostrar que el proyecto de la IA ha estado animado, desde sus orígenes, por una lógica antropomórfica: el deseo de que la máquina nos imite, piense como nosotros, encarne nuestros valores. Su tesis es deliberadamente contraintuitiva. Lejos de constituir la vía hacia una técnica responsable, la humanización de la máquina engendra confusiones conceptuales, sobreestimaciones técnicas y, sobre todo, distorsiones morales: cuando se atribuye al artefacto una interioridad que no posee (el *fantasma humano dentro de la máquina*, invirtiendo la vieja crítica de Ryle), se diluye la

frontera de la que depende la atribución correcta de la responsabilidad. La propuesta de De Miguel, que él denomina *deshumanización de la tecnología*, no es un gesto antihumanista sino su contrario: preservar la singularidad de lo humano (empatía, juicio ético, creatividad) precisamente marcando con nitidez qué es artefactual y qué no lo es. Los límites claros, escribe, hacen buenas fronteras.

Importa subrayar por qué este capítulo abre el eje y no lo cierra. La distinción que De Miguel defiende, entre regular y comprender la técnica sin proyectar sobre ella una agencia moral que no tiene, es la condición de inteligibilidad de todo lo que sigue. Si no se sabe qué clase de entidad es un sistema artificial, no se puede decidir si confiarle el cuidado de un enfermo, si responsabilizarlo del desplazamiento de un trabajador o si entregarle la custodia de la verdad pública. El error categorial, una vez cometido, contamina todas las preguntas aplicadas. De ahí que el primer deber ético en sociedades tecnificadas sea, paradójicamente, de orden ontológico: no confundir la herramienta con un igual.

### **La prueba del cuidado: confianza, dignidad y la tentación de delegar**

Establecida la distinción, el eje la somete de inmediato a su prueba más exigente. Isaac Carbajal y Pablo Ruiz Osuna examinan la interacción humano-máquina en la atención médica, un dominio donde la fragilidad es máxima y donde el



envejecimiento poblacional y la escasez crónica de cuidadores presionan hacia la introducción masiva de robots asistenciales. Su análisis recupera la ética del cuidado en su densidad antropológica: el cuidado no es un conjunto de tareas optimizables, sino una experiencia relacional entre dos personas, anclada en una dignidad ontológica que, sostienen con Bartoli y Postigo, no depende de la funcionalidad, la conciencia ni el reconocimiento social, sino del ser sustancial de la persona. De esta premisa se sigue su cautela: el robot puede aliviar, acompañar y monitorear, pero la empatía y la transformación mutua que constituyen el cuidado resisten la replicación mecánica.

El capítulo introduce además el concepto que se volverá decisivo en el resto del eje: la confianza. Recogiendo a O'Neill, Han, Harari e Innerarity, los autores muestran que confiar en un robot no es un cálculo sobre su eficacia técnica, sino una decisión cargada de implicaciones políticas: cuando una sociedad deposita el cuidado de sus miembros más vulnerables en sistemas automatizados, está tomando una decisión sobre el tipo de sociedad que quiere ser, y corre el riesgo de despolitizar, bajo la apariencia de neutralidad técnica, lo que son responsabilidades colectivas. La advertencia es fina: la confianza diferencial (robots para los pobres, cuidado humano para las élites) puede normalizar un sistema de cuidado estratificado. La conclusión es de prudencia: mientras no exista una ética de las máquinas respaldada por estudios de largo plazo, el robot debe ser

auxiliar y no sustituto, sujeto a supervisión humana significativa. La pregunta de De Miguel, qué significa preservar lo humano, recibe aquí su primera respuesta concreta: significa no delegar el reconocimiento.

## **Del cuidado al trabajo: la justicia distributiva de la automatización**

Leonardo Manríquez-Méndez desplaza la escala: del rostro singular del paciente a la estructura social del mercado laboral. Su revisión crítica de la evidencia empírica sobre el impacto de la IA en el empleo evita deliberadamente tanto el optimismo como el catastrofismo, y se sitúa en una posición que el propio capítulo califica de institucional. La IA, muestra, produce simultáneamente desplazamiento y creación: destruye empleos en múltiples niveles de cualificación, incluidos, contra la creencia común, los cognitivos y gerenciales medios, y a la vez genera roles nuevos en ciberseguridad, salud digital y análisis de riesgos. El efecto neto no está predeterminado por la tecnología; depende de las decisiones políticas y del diseño institucional, tesis que el autor toma de Acemoglu y Johnson.

Lo que vuelve este capítulo propiamente *ético*, y no meramente económico, es su insistencia, apoyada en Sen y en Arendt, en que el trabajo es fuente de dignidad, identidad y autorrealización, de modo que la exclusión laboral no es solo una pérdida de ingresos sino una herida a la condición de

sujeto. La automatización desregulada amenaza con producir lo que Harari llamó una *clase inútil*: no superflua para la producción, sino privada del reconocimiento que el trabajo confiere. La posición intermedia de este capítulo en el eje es estructuralmente significativa. Hereda de Carbajal y Ruiz Osuna la categoría de dignidad y la proyecta sobre la justicia distributiva; y prepara, con su énfasis en la institucionalidad, la pregunta que el capítulo siguiente convertirá en su centro: ¿cómo se legitiman, ante ciudadanos libres e iguales, las decisiones colectivas sobre la técnica?

### **La legitimidad epistémica: ¿por qué nadie debe custodiar la verdad?**

Este capítulo opera como bisagra normativa del eje. Su diagnóstico es que las democracias liberales atraviesan una crisis de legitimidad de un tipo específico: *epistémica*; distinta de las crisis de eficacia o de representación: lo que se erosiona no es esta o aquella creencia, sino la infraestructura compartida que permitía que la corrección de una falsedad fuese reconocida como tal por quien la sostiene. La posverdad, la desinformación algorítmica y la tribalización digital no operan tanto difundiendo proposiciones falsas cuanto socavando la confianza en los mecanismos sociales de validación del conocimiento. Frente a ello, el capítulo examina y rechaza la tentación que llama *perfeccionismo epistémico*: la pretensión de que el Estado, una casta de expertos o las plataformas asuman la custodia de la verdad

pública. Tal remedio, argumenta, es autodestructivo, porque pretende producir coercitivamente una confianza que la coerción no puede producir; quien cura la enfermedad matando al paciente no ha curado nada.

La alternativa, un *liberalismo epistémico*, es epistémicamente sensible pero no perfeccionista, y gira en torno a una distinción que ilumina retrospectivamente todo el eje: la que media entre proteger las condiciones del juicio y sustituir el juicio. El Estado liberal puede y debe intervenir sobre la arquitectura del espacio público: exigir transparencia algorítmica, dismantelar redes inauténticas, garantizar el pluralismo informativo; sin pronunciarse jamás sobre la verdad de los contenidos que en él circulan. Es la misma frontera que De Miguel trazaba entre regular la técnica y antropomorfizarla, ahora aplicada a la gobernanza de la esfera pública. Y es la misma lógica de la confianza que Carbajal y Ruiz Osuna situaban en el cuidado: la confianza pública es un bien relacional que ninguna autoridad fabrica, pero que las instituciones pueden merecer o defraudar. La gobernanza algorítmica de lo público aparece así como el caso límite donde el perfeccionismo se disfraza de moderación de contenidos, y donde el problema no es que los sistemas yerren sobre lo verdadero, sino que concentran un poder de configurar el espacio común de razones sin satisfacer condición alguna de legitimidad.

## ¿Quién diseña el futuro? La apertura geopolítica

Constanza Rivano da al eje su movimiento de mayor amplitud. Si los capítulos previos preguntaron qué es la máquina, cómo cuidar, cómo distribuir y cómo legitimar, su capítulo pregunta *quién imagina el futuro y desde qué posición epistémica lo hace*. El transhumanismo, sostiene, no es una mera especulación sino un diseño cultural del porvenir, sostenido por instituciones, corporaciones e inversión, que presenta como aspiración universal de la especie una visión nacida en ecosistemas tecnocientíficos muy localizados del Norte global. Apoyada en la crítica de la colonialidad (Quijano, Mignolo, Dussel) y en la ontología relacional del Buen Vivir (Acosta, Escobar), Rivano muestra que esa universalidad es una localización que se desconoce a sí misma, y que América Latina puede intervenir críticamente para reorientar el proyecto hacia la justicia estructural, la sostenibilidad ecológica, el reconocimiento de la vulnerabilidad y la deliberación democrática.

Su propuesta de un transhumanismo posliberal converge, desde la periferia, con el liberalismo epistémico que la precede: ambos desplazan el eje de la pregunta desde el qué (la corrección sustantiva del resultado) hacia el quién y el cómo (la autoridad legítima y el procedimiento). Y ambos recuerdan que mejorar la vida no equivale a abandonar la fragilidad, la dependencia y el vínculo con el mundo común, sino acaso a cuidarlos mejor. Es la misma vulnerabilidad que

Carbajal y Ruiz Osuna defendían como constitutiva de lo humano, ahora elevada a criterio civilizatorio. Con esa ampliación del *demos* (el futuro tecnológico no pertenece solo a quienes financian la investigación, sino a quienes habitarán sus consecuencias) Rivano deja planteada, sin resolverla, una última pregunta: bajo qué condiciones materiales ese futuro compartido es, siquiera, posible. Es esa pregunta la que recoge el capítulo que cierra el eje.

### **El límite del medio: eficiencia, suficiencia y los tiempos del antropoceno**

David Manolo Sailer cierra el eje con un movimiento que, a primera vista, parece cambiar de terreno, del diseño del futuro a la gestión de los recursos, y que sin embargo culmina la misma operación de desenmascaramiento con que el eje comenzó. Así como De Miguel mostraba que la humanización de la máquina no es un dato neutro sino una lógica con genealogía e implicaciones morales, Sailer muestra que la eficiencia, categoría que suele presentarse como métrica técnica y valorativamente muda, es en realidad un dispositivo de racionalidad ligado al proyecto moderno de dominio sobre la naturaleza. Su análisis del efecto rebote (*rebound effect*), el fenómeno por el cual las ganancias de eficiencia terminan siendo absorbidas, y a menudo superadas, por el incremento del consumo que ellas mismas habilitan, demuestra que optimizar los medios no basta para alcanzar el fin que se persigue: la reducción absoluta en el uso de recursos.

De ahí la propuesta central del capítulo, un desplazamiento desde la eficiencia de los medios hacia la resiliencia del fin, que reformula en clave ecológica la distinción que atraviesa todo el eje entre proteger las condiciones y sustituirlas, entre administrar una arquitectura y suplantarlo que esa arquitectura debía servir. Sailer llama a asumir una nueva honestidad que reconozca los límites de la mitigación e integre la adaptación, la resiliencia y la suficiencia como pilares de una ética climática robusta, capaz de afrontar sin ilusión las pérdidas que el Antropoceno vuelve ya inevitables. El eje se cierra así, no con una respuesta técnica sino con una advertencia normativa: ninguna arquitectura institucional, por sofisticada que sea su gobernanza del cuidado, del trabajo o de la verdad, puede sostenerse sobre un medio que se agota.

### **Convergencias, tensiones y un programa abierto**

Tres conceptos circulan entre los cinco capítulos y los suturan en un debate único. El primero es la confianza, entendida no como sentimiento sino como bien relacional irreducible al control: confianza en el robot cuidador, en las instituciones de validación del conocimiento, en quienes diseñan el futuro, y, en última instancia, en la posibilidad misma de que la acción presente no comprometa las condiciones de vida futuras. El segundo es la dignidad, afirmada como anterior a la funcionalidad y, por ello, no negociable ante la eficiencia técnica (ya se trate del paciente

en estado vulnerable, del trabajador desplazado, del ciudadano tratado como objeto de gestión epistémica o de las generaciones futuras tratadas como externalidad). El tercero es la frontera entre arquitectura y contenido, entre proteger las condiciones del ejercicio de la agencia y suplantarla: distinción que aparece, con notable consistencia, en la deshumanización de De Miguel, en la supervisión humana de Carbajal y Ruiz Osuna, en el diseño institucional de Manríquez-Méndez, en el test de doble condición de Fuentes Valdebenito, en la deliberación situada de Rivano y en el desplazamiento de la eficiencia de los medios hacia la resiliencia de los fines que propone Sailer.

Las divergencias son igualmente instructivas, y el volumen no las disimula. Donde De Miguel pide trazar con firmeza la línea humano-máquina para impedir toda atribución de agencia moral al artefacto, Carbajal y Ruiz Osuna admiten que ciertos autómatas podrían, bajo criterios estrictos, merecer alguna consideración moral; la frontera ontológica que el primero quiere nítida, los segundos la dejan porosa. Donde Fuentes Valdebenito defiende la objetividad de ciertos hechos como condición de la deliberación y rehúsa el relativismo, Rivano insiste en la pluralidad irreductible de los futuros y de las cosmologías; la tensión entre un universalismo de las condiciones epistémicas y un pluralismo de los horizontes civilizatorios queda productivamente abierta. Y mientras Manríquez-Méndez confía en la institucionalidad estatal como árbitro de la transición



tecnológica, las críticas a la colonialidad del poder que Rivano moviliza recuerdan que esas mismas instituciones han sido, históricamente, vehículos de subordinación, contraste que el capítulo de Sailer complica todavía más: si el propio paradigma de la eficiencia, presentado como neutro, resulta ser un legado del proyecto moderno de dominio, cabe preguntarse hasta qué punto las instituciones llamadas a gestionar la transición ecológica no reproducen la misma lógica que dicen corregir. Estas fricciones no son defectos de la compilación: son el mapa de las decisiones que una ética de las sociedades digitales, y de sus condiciones materiales, todavía debe tomar.

De ellas se desprende un programa de investigación que el eje deja deliberadamente abierto. ¿Cómo traducir institucionalmente la distinción entre arquitectura y contenido, sin que la regulación de la primera se deslice hacia la tutela del segundo? ¿Qué arquitecturas concretas de supervisión humana preservan la confianza en el cuidado mediado por máquinas a lo largo del tiempo, y no solo en el laboratorio? ¿Puede sostenerse una objetividad de los hechos que sea, a la vez, hospitalaria con la pluralidad de los marcos valorativos desde los que se los interpreta? ¿Cómo integrar una ética de la suficiencia y la resiliencia sin renunciar a los logros materiales que la eficiencia técnica ha hecho posibles? ¿Cómo financiar una educación cívica, un periodismo profesional y una ciencia abierta capaces de cultivar las virtudes epistémicas (veracidad, escucha, responsabilidad)

sobre las que descansa, en último término, todo el edificio? Estas preguntas exceden la filosofía y reclaman el concurso del derecho, la economía política, la epistemología social, la ecología política y el diseño institucional. Pero su formulación rigurosa es ya una conquista, y constituye la contribución de este eje al volumen.

Si algo enseña la lectura conjunta de estos trabajos es que la ética contemporánea ha cambiado de domicilio. No ha abandonado la relación interpersonal (el rostro del paciente, la palabra dada, el reconocimiento del otro siguen en su centro), pero ha descubierto que esa relación está hoy mediada por estructuras técnicas, y sostenida sobre recursos finitos, que la condicionan desde antes de que comience. Pensar moralmente exige, por ello, pensar las plataformas, los algoritmos, las arquitecturas informacionales y los límites materiales del planeta como lo que son: no instrumentos neutros ni fondos inagotables, sino configuraciones de poder que distribuyen agencia, vulnerabilidad y tiempo. El eje *Ética* de este volumen no pretende clausurar ese examen, sino ofrecer un vocabulario común y una invitación a proseguirlo con la seriedad conceptual que el problema merece. Que ese examen requiera, como sugieren al unísono sus autores, no solo instituciones astutas sino también ciudadanos lúcidos y virtuosos, es quizá su lección más exigente y, a la vez, la más esperanzadora.

# Sobre los límites del enfoque humanista en el desarrollo de la inteligencia artificial

*On the Limits of the Humanistic Approach in Artificial Intelligence Development*

**Pablo de Miguel<sup>2</sup>**

## **Resumen**

Este artículo examina críticamente los límites del enfoque humanista en el desarrollo de la inteligencia artificial (IA). Partiendo de una genealogía que conecta los autómatas antiguos con los sistemas algorítmicos contemporáneos, se analiza cómo la tendencia a antropomorfizar la tecnología ha orientado los debates filosóficos, políticos y técnicos actuales en torno a la IA. Si bien la alineación con valores humanos ha sido presentada como requisito esencial para un desarrollo

---

<sup>2</sup> Universidad de Salamanca. <https://orcid.org/0009-0007-6157-7005>

ético y responsable, se sostiene que dicha estrategia puede generar confusiones conceptuales, sobreestimaciones técnicas y distorsiones morales. Frente a ello, se argumenta a favor de una “deshumanización de la tecnología” entendida como la preservación de su carácter no humano, lo que permitiría reconocer tanto sus fortalezas específicas (capacidad de cómputo, objetividad, eficiencia) como las singularidades humanas (empatía, juicio ético, creatividad). La propuesta final es considerar la IA no como un espejo de la mente humana, sino como la posibilidad de inteligencias radicalmente distintas que amplíen el horizonte de lo pensable y refuercen la complementariedad entre técnica y humanidad.

**Palabras clave:** Inteligencia artificial; filosofía de la tecnología; antropomorfismo; ética; deshumanización de la tecnología.

## **Abstract**

This article critically examines the limits of the humanistic approach to artificial intelligence (AI) development. By tracing a genealogy from ancient automata to contemporary algorithmic systems, it explores how the anthropomorphization of technology has shaped current philosophical, political, and technical debates around AI. While alignment with human values has been promoted as a necessary condition for ethical and responsible AI, this

strategy often entails conceptual confusions, technical overestimations, and moral distortions. As an alternative, the paper argues for a “dehumanization of technology,” understood as preserving the non-human nature of machines. This perspective highlights the unique strengths of AI (computational capacity, efficiency, and objectivity) while safeguarding distinctively human qualities (such as empathy, ethical judgment, and creativity). Ultimately, the proposal is to view AI not as a reproduction of human intelligence but as the emergence of radically different forms of cognition that enrich our understanding of intelligence and foster a complementary relationship between humanity and technology.

**Keywords:** Artificial intelligence; philosophy of technology; anthropomorphism; ethics; dehumanization of technology.

## **Del autómatas al algoritmo**

La noción de crear autómatas o seres mecánicos humanoides no es en absoluto nueva; desde la antigüedad clásica abundan relatos sobre criaturas artificiales que evidencian un imaginario técnico temprano, un hilo simbólico que conecta aquellos mitos remotos con los actuales esfuerzos por imitar las facultades cognitivas humanas.

Distintas tradiciones culturales han albergado el profundo anhelo de reproducir la racionalidad humana

mediante artefactos, lo que sugiere que este deseo está lejos de ser una extravagancia moderna, pues tiene raíces milenarias en diversas culturas. Ya en la mitología griega aparecen autómatas capaces de imitar funciones humanas: el dios Hefesto forja sirvientes mecánicos con forma de mujeres dotadas de conocimiento e incluso construye al gigante de bronce Talos para proteger Creta, Aristóteles llegó a especular que, si cada herramienta pudiera operar por sí sola entonces los amos ya no necesitarían esclavos, incluso los ingenios de Herón de Alejandría describían estatuas móviles y teatros automáticos con mecanismos accionados por agua o vapor (Mayor, 2018, pp. 7-30). Lo que sugiere la admiración y, a la vez, el recelo ante invenciones que replican o superan las capacidades humanas se remontan a miles de años atrás, evidenciando que nuestros sueños y pesadillas tecnológicas no son exclusivos de la era moderna. Así, la tradición occidental ha oscilado entre utopías de creación la *techné* y temores a una *hybris* o usurpación de las prerrogativas divinas, que anticipan inquietudes análogas a las que hoy suscita la inteligencia artificial.

Este impulso imaginario de conferir cualidades humanas a lo inanimado no se limitó a la Antigüedad, en el siglo XVIII, en plena Ilustración, esta vieja aspiración toma forma en sofisticados autómatas de exhibición que intentaban replicar procesos de la vida y la inteligencia. El ingeniero francés Jacques de Vaucanson maravilló a Europa con su famoso *pato mecánico*, capaz de “comer” grano, digerirlo y

excretarlo, así como con figuras humanoides autómatas que tocaban la flauta o el tambor. Lejos de ser curiosidades aisladas, estos inventos fueron parte de un empeño ilustrado por demostrar que los procesos orgánicos y cognitivos podían replicarse mecánicamente. Igualmente, el célebre *Turco Ajedrecista* construido por Wolfgang von Kempelen pretendía encarnar un autómata pensante que jugase al ajedrez. Aunque en este último caso se trató de un fraude con un humano oculto en su interior, la fascinación que generó en el público refleja la convicción ilustrada de que incluso la mente podría objetivarse en una máquina. Inaugurando de esta forma una dinámica por mecanizar una función vital dentro de una máquina (Riskin, 2003). En este contexto no es casual que el filósofo materialista de La Mettrie, ya en 1747 titulara *L'homme machine* a su famoso tratado, afirmando explícitamente la continuidad esencial entre el hombre y el autómata (de La Mettrie, 1961, p. 101-102).

A estas alturas, pudiera parecer que la inteligencia artificial contemporánea no es sino heredera de un impulso secular que se remonta a la antigua Grecia. Por lo tanto, se puede afirmar que la IA se inscribe en la prolongada historia de la automatización representando la última fase de una sucesión histórica de iniciativas que arrancó hace mucho. Desde la Revolución Industrial, la fantasía de la máquina pensante se entrelaza con la mecanización del trabajo y del conocimiento. Por ejemplo, el inventor Charles Babbage, al concebir su máquina analítica en el siglo XIX, se inspiró

explícitamente en la división del trabajo fabril para externalizar cálculos mentales en un dispositivo automático. Expresiones históricas como esta, vendrían a confirmar que la IA moderna no surge ex nihilo, sino como la continuación de un proyecto largamente acariciado como es la idealización de seres artificiales iniciada en la antigüedad y desarrollada a lo largo de la historia, que culmina en la inteligencia artificial actual (Pasquinelli, 2023, p. 206-219). Lejos de suponer una ruptura radical, la IA de nuestros días puede verse entonces como la prolongación de un antiguo proyecto *antropotécnico*: el empeño humano por objetivar sus propias facultades cognitivas en soportes externos. Pero con un pequeño matiz, siempre ha sido un proyecto alimentado por el imaginario simbólico y los logros técnicos de cada época. Y es que si bien hay un constante por introducir en ingenios mecánicos cualidades humanas, lo que nunca ha permanecido constante en este largo recorrido es el paradigma teórico que servía en cada caso para enmarcar y evaluar las posibilidades del proyecto.

Cada era ha reinterpretado el sueño de la máquina humana a la luz de sus concepciones científicas dominantes, desde la mecánica clásica hasta la teoría de la computación de nuestros días, adaptando la vieja aspiración a los nuevos conocimientos tecnológicos. Con el desarrollo de las máquinas de Turing, y la plena comprensión de las implicaciones de la nueva teoría, surgió la convergencia entre el potencial técnico de los ordenadores y la posibilidad real de



reproducir artificialmente nuestras capacidades (Alonso, 2006).

Este punto resulta crucial para comprender la inflexión que quizás hoy se esté gestando. Reconocer de forma crítica que la IA es la continuación de un sueño antiguo de humanizar la técnica tiene importante implicación: nos invita a cuestionar la lógica antropomórfica que ha dominado históricamente este proyecto.

Durante décadas, el objetivo predominante ha sido lograr máquinas que pensaran y se comportaran como humanos, imitando fielmente nuestra conducta. De hecho, la propia fundación del campo de IA en 1956, con la conferencia de Dartmouth, se apoyó en la conjetura de que cada aspecto de la inteligencia humana puede ser descrito con tanta precisión que una máquina puede simularlo (McCarthy et al., 2006). No es sorprendente entonces que gran parte de la narrativa en torno a la IA haya enfatizado su semejanza con la mente humana. Sin embargo, tras este reconocimiento de continuidad histórica, surge con más fuerza la necesidad de reflexionar sobre los límites y el alcance de ese paradigma antropomórfico y explorar otras vías.

Precisamente, debido al objeto de estudio y a las necesidades del análisis el marco que ofrece la filosofía de tecnología -cuyo propósito principal es evaluar y analizar los sistemas técnicos, así como las actividades vinculadas a su

desarrollo- resulta el idóneo para desarrollar las tareas que nos proponemos en el presente texto. Pero para ello, lo primero que debemos hacer será analizar de qué manera y desde qué perspectivas filosófico-políticas contemporáneas se está abordando el fenómeno de la inteligencia artificial.

## **Tendencias filosófico-políticas actuales en torno a la Inteligencia Artificial**

El desarrollo actual de la tecnología ha abierto la posibilidad real de recrear, en agentes artificiales, ciertas cualidades cognitivas propias de los seres humanos. Frente a este avance, la percepción social de la inteligencia artificial oscila entre el reconocimiento de sus potenciales beneficios y la preocupación por sus posibles efectos no deseados (Maslej et al., 2024). Si bien se admite que la IA puede ser una herramienta idónea para alcanzar objetivos valiosos, existe también un rechazo a que sea la propia tecnología quien determine dichos fines. En este contexto, se ha consolidado un amplio consenso en los ámbitos político, académico y empresarial: los sistemas de inteligencia artificial deben alinearse con los valores humanos para garantizar que sus comportamientos y resultados generen un beneficio real para las personas.

Esta idea alineación de valores implicaría diseñar la IA de modo que actúe coherentemente con principios éticos fundamentales en todas las fases de su desarrollo y uso. El

objetivo no es otro que lograr que los sistemas de inteligencia artificial estén correctamente alineados con los valores humanos (Russell, 2019). En otras palabras, dotar a los agentes artificiales de las normas y formas de operar humanos es la condición necesaria para que la IA actúe de forma beneficiosa y responsable.

La importancia de este alineamiento derivaría de las potenciales consecuencias negativas de una IA no alineada con valores humanos. Si un sistema inteligente persigue objetivos que no reflejan consideraciones humanitarias, éticas en el fondo, podría generar decisiones contrarias al bienestar humano. Por ello, se aboga por orientar el desarrollo de la IA hacia el humanismo, evitando que la tecnología persiga metas contrarias o desvinculadas a nuestros valores. Debido a las dificultades de garantizar que los sistemas inteligentes actúen conforme a los valores humanos en todas las etapas, desde el diseño hasta su uso, se argumenta que la IA debe dirigirse a fines humanos (Han & Kelly, 2022). En los años recientes, múltiples iniciativas globales han convergido en torno a ciertos principios éticos fundamentales para una IA confiable: valores como la transparencia, la justicia y equidad, la no maleficencia, la responsabilidad y la privacidad se reconocen como ejes clave que deben incorporarse en los sistemas de IA (Jobin et al., 2019).

La necesidad de incluir en todas las etapas de la tecnología principios y valores sería la herramienta adecuada para integrar consideraciones de derechos humanos y bienestar social en el desarrollo algorítmico. Pasando de una visión puramente tecnocéntrica a una dimensión ética. Se enfatiza así que la IA debe diseñarse con una brújula moral que incorpore valores humanos para orientar su comportamiento y evitar consecuencias adversas para la sociedad.

Desde una perspectiva fenomenológica, se podría definir un conjunto finito de valores humanos aplicable a las tareas que realiza un sistema de inteligencia artificial y que, con la mediación tecnológica adecuada, podría fomentar la alineación de estos valores entre los diseñadores de la IA y sus usuarios. Por ejemplo: equidad en sistemas de reclutamiento, la seguridad y autonomía en vehículos autónomos, etc. Por lo que la idea de identificar y formalizar un conjunto de valores humanos fundamentales para guiar el diseño de la IA.

Además, para que esta tecnología transmita valores humanos el desarrollo de la IA debe integrar perspectivas de otras disciplinas humanistas, como son la ética y la estética (Sanguinetti, 2023). Ya que si se busca que los sistemas de IA actuales encajen con sentido en el mundo humano mediante herramientas provenientes de las artes y las humanidades. Cobra especial relevancia la ética en el nuevo humanismo adaptado a la era digital de tal forma que sirva como base para

el desarrollo de una inteligencia artificial responsable. Obligando a los sistemas a asumir un compromiso firme con los valores humanos para orientar su evolución.

Un argumento central dentro de esta dimensión ética es la prevención de sesgos algorítmicos y la discriminación. Existe evidencia concreta que ha mostrado como algoritmos comerciales de reconocimiento facial presentaban tendencias discriminatorias, con tasas de error significativamente mayores al identificar a mujeres con piel oscura en comparación con hombres de piel clara (Buolamwini & Gebru, 2018). De lo cual se deduce que si la IA no se desarrolla con sensibilidad hacia valores humanos universales, como son por ejemplo la equidad e inclusión, corre el riesgo de perpetuar y amplificar injusticias existentes.

Este tipo de sesgos derivan de datos o diseños que no incorporan diversidad ni valores de equidad, y han motivado llamados a una ética algorítmica más robusta. En la práctica, esto implica integrar activamente criterios de equidad, diversidad y respeto a los derechos en los conjuntos de datos, en los modelos y en las métricas de desempeño de la IA. Por lo tanto, el verdadero desafío no es solo acordar principios abstractos, sino traducirlos en requisitos técnicos y estándares implementables durante la construcción de sistemas inteligentes.

En este punto surge el paradigma del diseño centrado en el ser humano enfocado a cómo las tecnologías de IA deben concebirse con las necesidades, valores y limitaciones humanas como punto de partida, en lugar de forzar a las personas a adaptarse a la lógica de las máquinas (Shneiderman, 2020). En esta propuesta de IA centrada en el humano (HCAI por sus siglas en inglés *Human-Centered AI*), los sistemas inteligentes se diseñaron para apoyar distintas cualidades positivas de las personas (autosuficiencia, creatividad, etc.), a la vez que se otorga al usuario comprensión y control sobre las acciones del sistema, en lugar de operar como cajas negras opacas e inescrutables. Invirtiendo la prioridad tradicional, primero las personas y después la tecnología, la IA tendría la capacidad de potenciar a las personas en lugar de intentar simplemente emularlos o reemplazarlos.

En lugar de aspirar a máquinas totalmente autónomas que traten a los usuarios como operadores pasivos, se busca diseñar sistemas como colaboradores o herramientas que amplifiquen las capacidades humanas. Un cambio de paradigma que pasaría de intentar emular la inteligencia humana a empoderar a las personas. El objetivo sería lograr sistemas tecnológicos avanzados que no anulen la agencia humana, sino que la complementen. En la práctica, esto se traduce en interfaces que ofrecen opciones y transparencia sobre el estado y las acciones de la IA, así como mecanismos

de retroalimentación donde el usuario pueda corregir o ajustar el comportamiento del sistema.

La necesidad de humanizar la tecnología también se manifiesta en desarrollos técnicos concretos destinados a incorporar las consideraciones humanas en los sistemas de IA. Uno de los ejes técnicos más destacados es asegurar la supervisión y el control humanos en los bucles de toma de decisión automatizada. En contraste con la noción de una IA totalmente autónoma, la investigación reciente enfatiza arquitecturas que mantienen a las personas involucradas en las decisiones críticas del sistema, esto resulta crucial en entornos de alto riesgo. Una estrategia técnica recurrente para ello consiste en incorporar valores durante el entrenamiento de los modelos mediante enfoques como el aprendizaje por refuerzo con retroalimentación humana (Reinforcement Learning from Human Feedback, RLHF). Por ejemplo, RLHF se ha utilizado para ajustar modelos de lenguaje de gran escala a las preferencias humanas en cuanto a la corrección y adecuación de sus respuestas (Ouyang et al., 2022). De modo que el sistema no solamente optimiza una métrica numérica, sino también criterios alineados con lo que los humanos consideran aceptable. Junto con RLHF, se emplean técnicas de diseño de recompensas seguras, aprendizaje inverso de preferencias humanas, y pruebas de estrés ético, entre otras.

Otra línea técnica fundamental es el desarrollo de IA explicable y transparente (XAI, denominado así por sus siglas en inglés, *Explainable Artificial Intelligence*). Una queja recurrente sobre los modelos de aprendizaje profundo modernos es su carácter de “caja negra”, que dificulta comprender por qué la máquina tomó una determinada decisión. La explicabilidad ha dejado de ser una demanda exclusiva de desarrolladores para convertirse en un requisito esencial que el ciudadano lego entienda y confíe en la IA, especialmente en dominios sensibles como salud, finanzas o justicia (Liao & Varshney, 2022). Al ser la explicabilidad una propiedad intrínsecamente centrada en lo humano, pues son las personas quienes necesitan entender, se ha adoptado enfoques de XAI centrados en el usuario, integrando conocimientos de interacción humano-computadora y diseño de experiencia de usuario para crear explicaciones útiles y adaptadas a las necesidades reales de los usuarios. La importancia de esta dimensión técnica radica en que, sin comprensibilidad, los usuarios no pueden ejercer juicio ni dar retroalimentación informada sobre las decisiones de la IA. En respuesta a estos desafíos, se han desarrollado técnicas como modelos explicativos simplificados, visualizaciones del razonamiento, auditorías algorítmicas y resúmenes de decisiones, todo orientado a brindar mayor transparencia.

En suma, dentro de tendencias filosófico-políticas actuales en torno a la inteligencia artificial sobresale la investigación en alineación de la IA con valores humanos, que



explora cómo inculcar de forma efectiva normas y técnicas de preferencias humanas en algoritmos autónomos.

## **Crítica al enfoque humanista del desarrollo de la Inteligencia Artificial**

Como se puede apreciar en los debates actuales sobre inteligencia artificial existe una fuerte tendencia a hacer converger, en esencia, la tecnología con cualidades humanas. Se empuja a la tecnología a buscar fagocitar las cualidades humanas, no solo para imitarlas, sino para integrarlas en su propio ser. Esta perspectiva humanista presupone que incorporando rasgos humanos a las máquinas se facilita la supresión de comportamientos perjudiciales. Puesto que para hacer la IA beneficiosa para la sociedad debemos, de algún modo, dotarla de estas propiedades como empatía, responsabilidad o incluso cierto sentido de moralidad.

Esta tendencia responde a los continuos problemas que actualmente plantea la tecnología y que cuentan con un claro interés filosófico y que se inserta en los modelos tradicionales de lo que se considera que es objeto de reflexión filosófica. Siguiendo estas mismas pautas cabe preguntarse de forma crítica el sentido de esta estrategia y si realmente se alinea con los objetivos que se propone o si por el contrario encierra confusiones conceptuales importantes. Todo ello, mediante un juicio razonado, sopesando ventajas e

inconvenientes, así como sus condiciones de posibilidad y las alternativas, si es que las hubiera.

No se trata de discutir si las máquinas pueden alcanzar o no ciertos atributos semejantes a los humanos como, inteligencia o conciencia, sino reflexionar sobre la necesidad y utilidad de exigirles atributos humanos para que sean beneficiosas.

De la misma manera que hemos empezado este texto describiendo la lógica antropomórfica en la que se inserta este proyecto, será la primera sometida a examen. Y es que de ese proyecto antropoformista está conduciendo, tanto a tecnólogos como a filósofos, a una obsesión por hacer que las máquinas sean como nosotros. Sin embargo, vale notar que precisamente la falta de cualidades humanas es lo que da valor a las máquinas (Innerarity, 2025, pp. 246-249), las máquinas no se cansan, no sienten hambre ni tienen emociones que nublen su juicio. Es precisamente la diferencia de sus propiedades frente a las nuestras lo que hace posible que nos brinden servicios tan valiosos. Los sistemas de IA pueden trabajar con mayor rapidez y fiabilidad que los humanos porque no se distraen con fatiga o impulsos fisiológicos; esa ausencia de factores humanos (cansancio, sesgos, emociones, etc.) lo que puede hacer a las máquinas eficientes y útiles para nosotros (Heilingner, 2022).

Siguiendo la misma estela de crítica a la lógica antropomorfismo, la investigadora Adriana Placani (2024) lo analiza como una exageración y falacia a la vez. Por un lado, al atribuir rasgos humanos a sistemas que no los poseen, tendemos a exagerar sus capacidades. Por otro lado, ese mismo gesto conduce a una falacia moral al distorsionar nuestros juicios sobre la máquina, sobre su carácter moral o responsabilidad. En el campo de la ética esta visión acarrea consecuencias negativas, ya que articula el debate en torno a la IA como si esta fuera un agente moral.

Se asigna a un ente de naturaleza totalmente distinta atributos que, hasta donde sabemos, pertenecen a otro orden del ser humano. La problemática en este punto reside en a partir de esas supuestas cualidades se pueden acabar emitiendo juicios morales inapropiados. De un parecido superficial o un deseo de humanizar, se puede llegar a asumir una propiedad ontológica que no está realmente presente. Esta exposición no se limita a un mero caso teórico; de hecho, ya se han dado debates públicos sobre los derechos de los robots. En 2017 el Parlamento Europeo contempló en un informe la idea de conceder a los sistemas de IA más avanzados un estatus de “persona electrónica” para delimitar responsabilidades legales (Parlamento Europeo, 2017). El filósofo Gilbert Simondon advertía ya en la segunda mitad del siglo pasado contra lo que llamaba “humanismo fácil” (Simondon, 2013, pp. 31-33). Este tipo de humanismo consistiría en atribuir un valor absoluto a lo humano frente a

lo técnico, o peor aún, en personalizar la técnica sin comprenderla verdaderamente. Simondon proponía, en cambio, entender los objetos técnicos en su génesis y en su relación con la humanidad, sin caer en animismos ni en simplificaciones dualistas. Reflexionar sobre la relación humano-tecnología no implica ignorar los límites de nuestra condición, nuestro humanismo debe reconocer su posición respecto a la técnica, incluyendo nuestras deficiencias. Un humanismo tecnológico sensato no consiste en difuminar la línea entre humanos y máquinas, sino en marcarla con mayor claridad para luego gestionarla sabiamente.

Es similar, aunque invertido, a la crítica del filósofo Gilbert Ryle al “fantasma en la máquina”, donde Ryle denunciaba la idea de un alma o mente inmaterial dentro del organismo humano como un error categorial (Ryle, 2009, pp. 5-8). En el caso de la IA, el error sería poner un fantasma humano dentro de la máquina tecnológica, es decir, proyectarle una interioridad que no tiene.

El antropomorfismo alimenta esas proyecciones seductoras a futuro: la fantasía de una IA “humana” plena. Pero existen razones para sostener que los debates sobre escenarios futuristas no siempre son acompañados por un debate racional y es necesario buscar una visión más precisa de la tecnología que no esté alimentada por expectativas ni temores. Como hemos visto la importancia del modo en que concebimos la IA, especialmente si la antropomorfización

ingenuamente tiene consecuencias prácticas en el ámbito moral. De la misma manera también ocurren distorsiones epistemológicas notables, como es la pérdida de perspectiva sobre el origen y propósito de la tecnología. Esta tendencia desplaza la discusión hacia escenarios futuristas poco tangibles, IAs conscientes o robot-humanoides a costa de ignorar problemas reales.

Si no definimos cuidadosamente el marco en que queremos que ocurran los debates corremos el riesgo de perdernos en discusiones estériles o, peor aún, de diseñar políticas públicas basadas en nociones equivocadas. Por lo tanto, lejos de pujar por una humanización de la tecnología hay que defender la falta de humanidad de las mismas.

Desde luego, adoptar esta posición no significa reconocer que los sistemas de inteligencia artificial carezcan de problemas o abrazar el tecno-optimismo donde simplemente habría que dejar a la técnica desarrollarse. En el desarrollo tecnológico influyen factores internos, relacionados utilizar los medios técnicos más adecuados para los lograr los fines propuestas, y externos, que se relacionan con la valoración del sistema tecnológico para la sociedad en su conjunto. Es en este último punto de vista donde se vuelcan los criterios para una evaluación tecnológica, donde la sociedad actúa siempre como filtro para la innovación. No todo diseño tecnológico es valioso para la sociedad y puede no

ser adoptado si se perciben consecuencias negativas o contrarias a los valores imperantes.

Atender a las complejidades del desarrollo tecnológico y a la interacción entre los criterios de evaluación, ya sean internos o externos, y la sociedad como filtro, permite trazar un camino para prevenir las consecuencias negativas de dicho desarrollo sin necesidad de formular una nueva teoría humanista sobre la inteligencia artificial. Además, por contradictorio que parezca, la filosofía y la ética pueden tener más peso de esta forma frente a las teorías filosóficas que se limitan a establecer principios. Ya que las grandes oportunidades que ofrece el desarrollo tecnológico siempre se combinan con importantes desafíos, que de ser pasados por alto pueden generar un rechazo social. Y la reflexión ética, sobre lo que se considera valioso o no, supone una herramienta valiosa para los tecnólogos y los decisores de políticas públicas con el fin de no desaprovechar las posibilidades que ofrece el desarrollo tecnológico ni permitir impactos negativos.

### **Argumentos a favor de la deshumanización de la tecnología**

Estas tendencias técnicas, filosóficas y políticas se empeñan en humanizar la inteligencia artificial están más basadas en lugares comunes y tópicos filosóficos que en juicios críticos razonados. Frente a ello, se puede establecer

una línea teórica a favor de una deshumanización de la tecnología. Lejos de proponer una visión anti-humana, esta idea busca en realidad salvaguardar tanto la eficacia de las máquinas como los valores humanos. La deshumanización de la tecnología, entendida no en un sentido peyorativo sino como la renuncia consciente a proyectar cualidades humanas en las máquinas, es una propuesta que busca proteger la ausencia de humanidad de las máquinas.

Defender la deshumanización de la tecnología implica reconocer que los sistemas artificiales, por muy sofisticados que sean, su principal fortaleza y contribución a la humanidad es carecer de conciencia, emociones y genuina comprensión humana. Al no humanizar la tecnología, se evita confundir su naturaleza, delimitando sus fortalezas propias (capacidad de cómputo, velocidad de procesamiento, análisis de datos) y las nuestras, (empatía, juicio ético, creatividad), de modo que cada parte cumple su rol de manera complementaria.

En este sentido, la conveniencia de trazar un límite nítido entre los seres humanos y las máquinas inteligentes resulta evidente. La IA no debe ser considerada ni tratada como un agente moral o un igual, pues carece de las condiciones ontológicas necesarias para asumir deberes y responsabilidades, los sistemas de IA son artefactos diseñados por humanos, no entidades autónomas con derechos propios. Fomentar la ilusión de que la IA es humana explota la ignorancia del público y confunde la

responsabilidad moral. Por lo tanto, mantener la distinción humano-máquina protege, por tanto, a la sociedad de malentendidos y asegura que la responsabilidad ética recaiga en quienes corresponde: los diseñadores, programadores y operadores humanos.

De igual modo, antropomorfizar la IA puede distorsionar nuestra percepción de sus capacidades y límites. Presentar a la IA en términos humanizados restringe nuestra comprensión a categorías humanas, induciendo comparaciones engañosas y sobrevalorando sus capacidades reales (Barrow, 2024). Igualmente, conduce a establecer sistemas de medición inadecuados ya que, como se ha dicho anteriormente, la falta de cualidades humanas es lo que da valor a las máquinas. Sin embargo, dentro de la lógica antropomórfica se constituye a lo humano como *benchmark* de lo no humano.

En la misma línea argumentativa se pueden resaltar los beneficios de la eficiencia amoral de los sistemas técnicos. En contextos de gran sensibilidad moral, la carencia de emociones puede ser una ventaja, la falta de subjetividad personal en la IA puede aportar consistencia y objetividad. Se valora, así, la eficiencia amoral de estas máquinas, cuya ejecución rigurosa de objetivos programados, sin dilemas emocionales, puede maximizar resultados beneficiosos de manera fiable.



De la misma forma, es importante cuestionar los mitos tecnológicos asociados a la IA, como el de la suplantación o el de las máquinas pensantes. Tales mitos responden más a prejuicios irracionales que a análisis rigurosos. La filosofía de la tecnología puede ayudarnos a desentrañar estas ilusiones y a comprender que la tecnología es, ante todo, una oportunidad de desarrollo humano y un motivo de esperanza, no de temor.

## **Conclusión**

Deshumanizar la tecnología no significa degradarla ni minusvalorarla, sino criticar el desarrollo humanista de la inteligencia artificial en relación a sí mismo, es decir, comparándolo con las posibilidades que esa misma teoría ofrece y situarla en su lugar propio: el de una herramienta poderosa al servicio de fines humanos. Solo manteniendo clara la frontera entre lo humano y lo artificial podremos aprovechar sus beneficios sin poner en riesgo nuestra dignidad, nuestras responsabilidades y nuestras relaciones genuinamente humanas.

Abogar por la deshumanización de la tecnología no significa demonizarla ni impedir mejoras en su interacción con nosotros, sino recordarnos constantemente su verdadera naturaleza no humana. Solo así podemos relacionarnos con ella con lucidez, evitando el humanismo ingenuo que difumina los límites entre artefactos y personas. La filosofía y

las humanidades tienen aquí un papel decisivo, no para antropomorfizar a la IA, sino para pensar las mejores formas de integrar sus innovaciones en la esfera social, sometiéndolas al filtro de la evaluación ética y política.

La proliferación de marcos éticos y guías de principios para la IA, como los promovidos por la Unión Europea, muestra la creciente preocupación por orientar la innovación tecnológica hacia valores compartidos. Sin embargo, como advierte Coeckelbergh (2020, pp. 60-62), el riesgo de una ética de *checklist* trivializa los dilemas de fondo, reduciendo la reflexión filosófica a un ejercicio de *compliance* superficial. Evitar esa tentación exige mantener vivo el debate moral y no delegarlo en listas estandarizadas de principios. El trabajo interdisciplinar entre tecnólogos y filósofos debe expandirse mucho más allá de estas prácticas, se deben buscar metodologías multidisciplinares reales.

Independientemente de ello, puede que la verdadera promesa de la IA no está en reproducir la mente humana, sino en permitir la emergencia de formas de inteligencia radicalmente distintas. Existen reflexiones que resultan especialmente iluminadoras sobre la agencias artificiales sin inteligencia que sugieren que, en lugar de ampliar el concepto de inteligencia para abarcar a la IA, conviene reconocer nuevas formas de agencia, no necesariamente dotadas de cognición, intención o estados mentales, que permitan entender la IA en su singularidad (Floridi, 2024).

Asumir esta perspectiva supone abandonar el espejismo narcisista de la mimesis y explorar modalidades no humanas de cognición, que podrían ampliar el horizonte de lo pensable y enriquecer nuestro entendimiento de la inteligencia misma. El desafío no es hacer que las máquinas piensen como nosotros, sino concebir otras inteligencias, genuinamente no humanas.

Este planteamiento entra de lleno con la cuestión abierta sobre la posibilidad de alcanzar una inteligencia artificial general, es decir, una máquina con inteligencia similar a la humana. El debate técnico-científico actual oscila entre quienes confían en el aprendizaje profundo y quienes reclaman nuevos paradigmas, pero, en cualquier caso, la pregunta clave es si conviene seguir persiguiendo el viejo sueño de la imitación humana o apostar por inteligencias inéditas. La respuesta a esta encrucijada marcará el rumbo del proyecto de la IA en las próximas décadas.

En conclusión, beneficiarnos de las capacidades computacionales de la IA exige mantener firmes nuestras propias responsabilidades y singularidades. Tecnología y humanidad pueden coexistir de forma complementaria, cada cual brillando en sus dominios. Preservar la falta de humanidad de las máquinas, paradójicamente, es una manera de proteger lo más valioso de la humanidad misma en la era digital: nuestra agencia, nuestra empatía y nuestra capacidad de comprender el mundo sin espejismos. Como se ha

argumentado, los límites claros hacen buenas fronteras: reconocer qué es humano y qué es meramente artefactual sienta la base para un futuro en el que los ingenios artificial siga siendo una extraordinaria herramienta al servicio de la sociedad, y no un reflejo distorsionado que nos extravíe de nuestra propia esencia.

## Referencias

- Alonso, E. (2006). De la computabilidad a la hipercomputación. *Azafea: Revista de Filosofía*, 8, 121–146.
- Barrow, N. (2024). Anthropomorphism and AI hype. *AI and Ethics*, 4, 707–711.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. En *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77–91).
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- De La Mettrie, J. O. (1961). *El hombre máquina*. Universitaria de Buenos Aires.
- Floridi, L. (2024). *AI as agency without intelligence: On artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis*. SSRN. <https://doi.org/10.2139/ssrn.5135645>
- Han, S., & Kelly, E. (2022). Aligning artificial intelligence with human values: Reflections from a phenomenological

- perspective. *AI & Society*, 37(4), 1383–1395.  
<https://doi.org/10.1007/s00146-021-01247-4>
- Heilinger, J. C. (2022). The ethics of AI ethics: A constructive critique. *Philosophy & Technology*, 35, 61.  
<https://doi.org/10.1007/s13347-022-00557-9>
- Innerarity. (2025). *Teoría crítica de la inteligencia artificial*. Galaxia Gutenberg.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.  
<https://doi.org/10.1038/s42256-019-0088-2>
- Liao, Q. V., & Varshney, K. R. (2021). *Human-centered explainable AI (XAI): From algorithms to user experiences*. arXiv.  
<https://doi.org/10.48550/arXiv.2110.10790>
- Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., & Clark, J. (2024). *Artificial intelligence index report 2024*. arXiv.  
<https://doi.org/10.48550/arXiv.2405.19522>
- Mayor, A. (2018). *Gods and robots: Myths, machines, and ancient dreams of technology*. Princeton University Press.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31,

1955. *AI Magazine*, 27(4), 12.  
<https://doi.org/10.1609/aimag.v27i4.1904>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Aspell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35.
- Parlamento Europeo. (2017, 16 de febrero). *Resolución del Parlamento Europeo con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL))*. Diario Oficial de la Unión Europea.  
[https://oeil.europarl.europa.eu/oeil/fr/procedure-file?reference=2015/2103\(INL\)](https://oeil.europarl.europa.eu/oeil/fr/procedure-file?reference=2015/2103(INL))
- Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence*. Verso.
- Placani, A. (2024). Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*, 4, 691–698.
- Riskin, J. (2003). The defecating duck, or, the ambiguous origins of artificial life. *Critical Inquiry*, 29(4), 599–633.
- Russell, S. J. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.
- Ryle, G. (2009). *The concept of mind: 60th anniversary edition*. Routledge.

- Sanguinetti, P. (2023). *Tecnohumanismo: Por un diseño narrativo y estético de la inteligencia artificial*. La Huerta Grande.
- Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 12(3), 109–124.
- Simondon, G. (2013). *El modo de existencia de los objetos técnicos*. Prometeo.

# Can We Trust Robots? A Brief Consideration of Human–Machine Interaction in Medical Care

*¿Podemos confiar en los robots? Una breve consideración sobre la interacción humano–máquina en la atención médica*

**Isaac Carbajal<sup>3</sup>; Pablo Ruiz Osuna<sup>4</sup>**

## **Abstract**

The global decline in birth rates and prolonged aging, combined with chronic shortages of human and material resources in the medical field, call for a rethinking of health care delivery worldwide. Even so, the principles of the ethics

---

<sup>3</sup> Universidad Católica del Maule. <https://orcid.org/0009-0001-7022-4822>

<sup>4</sup> Universitat Rovira i Virgili. <https://orcid.org/0000-0001-5510-8981>



of care—patient autonomy and dignity—must be upheld. Robotics and AI are poised to take on an expanding role; however, uncertainty persists regarding their ability to perform tasks that are traditionally reserved for humans. Beyond technical limitations, the large-scale deployment of these technologies raises profound ethical and philosophical dilemmas that remain largely unaddressed by the literature. Simple robots inspire clear judgments, but the emergence of genuinely intelligent automatons blurs the question of trust, which depends heavily on how each user conceptualizes the machine. Patients can also become attached to these devices, complicating their responsibility. For the moment, robots should serve as sophisticated assistants while a comprehensive machine-ethics framework—bolstered by long-term studies on sustained human-robot interaction—grows mature enough to foster genuine trust in clinical practice.

**Keywords:** Robot trust; human–machine interaction; medical robotics; care ethics; clinical artificial intelligence.

## **Resumen**

El descenso global de la natalidad y el envejecimiento prolongado, combinados con la escasez crónica de recursos humanos y materiales en la medicina, exigen repensar la prestación sanitaria a nivel mundial. Aun así, los principios de la ética del cuidado —autonomía y dignidad del paciente—

deben mantenerse. La robótica y la IA están llamadas a desempeñar un papel creciente, pero persisten dudas sobre su capacidad para asumir tareas tradicionalmente humanas. Más allá de limitaciones técnicas, la implantación masiva de estas tecnologías plantea dilemas éticos y filosóficos aún inexplorados. Mientras los robots simples generan juicios claros, la llegada de autómatas verdaderamente inteligentes oscurece la cuestión de la confianza, que depende mucho de cómo cada usuario conciba al robot. Además, los pacientes pueden encariñarse de estas máquinas, complicando las responsabilidades. Por ahora, los robots deberían actuar como auxiliares, mientras se desarrolla una ética de las máquinas respaldada por estudios de largo plazo.

**Palabras clave:** Confianza en robots; interacción humano-máquina; robótica médica; ética del cuidado; inteligencia artificial clínica.

## **Introduction**

Society has changed considerably in recent years. The structural shifts in demographic patterns are alarming and require urgent attention. Worldwide, the number of people aged 60 and over grew from 400 million in the 1950s to 700 million in the 1990s (Alonso Galbán et al., 2007). Simultaneously, a decline in younger cohorts has accelerated population ageing (Ismail et al., 2021). Although this phenomenon is most pronounced in countries such as Japan,

Spain, and Malta, it has begun to take hold in many other parts of the world, albeit at varying speeds. For instance, Malaysia's older adult population (those aged 60 and above) exceeded 11 percent in 2020 (Ismail et al., 2021).

Simultaneously, changes in household habits and family formation are evident, as there are more single-parent families or couples with only one child (Ismail et al., 2021). This situation raises a variety of challenges because the resources devoted to caring for older family members become an almost unbearable burden when parents have only one child.

Moreover, as has long been cautioned, "the world's population will soon reach its peak and begin to decline—and once the decline starts, it will never end" (Bricker & Ibbitson, 2019, p. 320). Spain offers a telling example: its birth rate has fallen continuously since 2009, reaching just 320,656 births in 2023, 78,330 of which were to non-Spanish mothers (INE, 2025). Although Spain's situation is especially worrisome, the global picture is similar: the worldwide birth rate dropped from 22 per thousand inhabitants in 2000 to 18.1 in 2020 (IndexMundi, 2020). For the moment, these figures are less alarming than they might be because medical advances have sharply reduced mortality; in 2020 the crude death rate stood at 7.7 per thousand (IndexMundi, 2020). However, the global population growth rate is edging downward, and at this pace, it could stagnate if current trends persist (IndexMundi,

2020). This suggests that the national economy and productivity may stall.

In just 50 years, the Latin America and Caribbean (LAC) region has undergone an ageing process comparable to what took Europe two centuries. By 2100, the number of older adults in LAC is expected to surpass that in Europe. In recent decades, mortality rates have decreased with increasingly advanced ages (Kallestrup-Lamb et al., 2024). Long-term care constitutes a major component of overall healthcare expenditure (Kallestrup-Lamb et al., 2024). In the United States, people aged 65 and older accounted for 30 % of healthcare spending in 2008, a figure projected to rise to 50 % by 2030 (Hajizadeh et al., 2025, p. 52).

In this regard, “some authors have concluded that loneliness can worsen quality of life, cause suffering, and be considered a risk factor for negative health outcomes, including mortality” (Molas-Tuneu et al., 2023). This issue poses particular challenges when we are talking about people with dementia and other mental illnesses (Deusdad, 2024).

This study aimed to address the ethical questions raised by the use of social and companion robots in older adults, children, and people with dementia. We will also examine issues surrounding human–machine interaction, such as the anthropomorphization of robots, whether autonomous agents can and should be trusted, and the long-

term consequences of delegating medical tasks to robotics. Our proposal stems from the current and foreseeable difficulties associated with a robot's ability to provide comprehensive and autonomous patient care. It underscores the need for an ethical debate concerning the opportunities and limits of deploying assistive robots within healthcare systems. Such technologies should support medical practice without entailing the replacement of healthcare professionals or undermining patient autonomy.

### **Ethics of Care, with Special Reference to People with Disabilities**

Throughout the human life cycle—from conception to death—people require care to varying degrees; therefore, dependency is an inherent dimension of human condition (Kittay et al., 2005). This fact becomes most evident during times of heightened vulnerability, such as the earliest and latest stages of life and during episodes of illness, when frailty or loss of autonomy becomes more pronounced. In some countries, this situation has become particularly complex owing to rising life expectancy, as noted in the introduction.

Care is so crucial to the human species that it is indispensable not only for social development but also for survival (van Nistelrooij & Visse, 2019). Within this paradigm, relationships between people—the existence of an

“I” and a “you”—and interdependence are fundamental to the ethics of care (Bevilacqua, 2016).

In some countries, the shortage of caregivers and the heavy burden they bear have prompted the introduction of social assistance robots in hospitals, schools, private homes, and elder-care facilities, among other settings. Among the many roles assigned to these robots, companionship is one of the most prominent (Aymerich-Franch & Ferrer, 2021).

In the context of care, various questions arise regarding the use of robots, which are explored in this section of the paper. One concept that has gained particular prominence in the debate is autonomy, as it often determines the intensity and type of care a person requires.

Kant (2002) defines the autonomy of the will as “that mode of the will by which it is a law unto itself, independently of any modality of the objects of willing” (p. 157).

This means that a person acts autonomously when guided by rational principles that they recognize as universal rather than by the external consequences of their actions. Similarly, in principlism-based bioethics, autonomy is defined as the capacity to determine what will be done with oneself and one’s own body, a capacity that presupposes an individual’s competence to act. Although this view has many

critics for lacking a social dimension, some scholars advocate a relational conception of autonomy (Varkey, 2021).

In the healthcare field, among the wide range of existing diseases, those that cause the greatest loss of independence and autonomy in older adults are Alzheimer's disease, cerebrovascular accidents (strokes), and Parkinson's disease (Maresova et al., 2019). Understanding and factoring these epidemiological aspects into any analysis is crucial because people who require care tend to share certain characteristics shaped by their disease experience and progression.

Several questions arise regarding the specific analysis of autonomy and decision-making capacity. For instance, what happens to people who cannot accept the presence of assistive robots in their care because of their mental health condition? In the future, should this preference perhaps be recorded in a document prepared in advance while individuals are still competent—something akin to today's advance directives?

In the same vein, this loss of autonomy is linked to vulnerability, which is defined as a condition in which a person is exposed to being harmed, influenced, or attacked and is also associated with fragility, weakness, and precariousness (Ten Have, 2016).

Considering all the factors discussed, care is difficult to define precisely. It is an activity, an attitude, and a commitment that incorporates personal experience and functions as a social process encompassing love, affection, and work (Busquets, 2019). At its core, care is experienced by two people —the care recipient and caregiver— highlighting both the complexity of the experience and the depth of its meaning. This insight is crucial because care is enriched by the experiences of both individuals involved.

Care ethics also acknowledges the interplay between a given context and values; in other words, situated persons create situated values (Pettersen, 2011). This means that, in the field of care, universal rules are not enough: various factors concerning the care recipient must be considered—such as culture, emotions, history, experiences, and current as well as overall condition, among others. From this context emerge the values that guide and shape caregiving.

Regarding caregiving practice, it is important to emphasise its dynamic nature: caregivers constantly face diverse situations, which obliges them to develop a range of skills for day-to-day life (Leocadie et al., 2020). These skills are not confined to technical tasks—such as providing hygiene and comfort, feeding, or mobilising—but also encompass the ethical and relational dimensions of care. This is because the care recipient's condition is inherently fluid—biologically, psychologically, and spiritually—making flexibility an



essential requirement. Caregivers must also know how to prioritise the needs of the person in their charge; in practice, the absence of such prioritisation is reflected in high hospital-readmission rates and the use of emergency services for non-urgent situations (Rodrigues Dixe et al., 2019). All of this points to a lack of information and judgement, affecting not only patients but also institutions and the wider community.

These situations are not exceptional; they are an everyday part of life and are embedded in family relationships and social coexistence. Therefore, addressing them must also be intersectoral.

Among the anthropological components of care, the concept of personhood stands out. Care is often understood through this lens, and the ethical dimension is clearly shaped by its anthropological framework. In this regard:

Human nature is that of a corporeal-spiritual reality. We are not merely spiritual or rational beings, nor purely material ones; although we are corporeal beings, our entire corporeality is shaped and animated by a spiritual principle that makes us what we are. (Bartoli Presas, 2020, p. 92)

From this corporeal-spiritual reality, the person derives an ontological dignity. It is defined as follows:

“Personal dignity does not depend on subjective perception, culture, prevailing opinion, or the social majority; rather, it is grounded in the very freedom of the human being” (Bartoli Presas, 2020, p. 112).

Accordingly, dignity is not an attribute or value bestowed by functionality or social recognition; it is rooted in the person’s very substantial being. Therefore, any attempt to reduce a human being to a single dimension objectifies her at an instrumental level, failing to recognise that she is an end in herself and cannot be used merely as a means.

Understanding this approach is essential, especially in clinical settings where conditions such as the persistent vegetative state can be relatively common in certain departments.

Thus, it is asserted that ontological dignity endures even under conditions of extreme vulnerability. This is hard to grasp today, when certain schools of thought grant dignity only to those capable of exercising their higher faculties—consciousness, autonomy, will, and rationality.

This analysis is valuable in an environment of rapid technological growth, where progress often solidifies before reflections aimed at safeguarding human dignity can take place.

The philosopher Elena Postigo warns of this risk: “The development of instrumental-technical reason is advancing very rapidly, whereas sapiential reason moves slowly, unable to analyse calmly and in detail all that these changes entail” (Postigo Solana, 2022).

It seems that attitudes toward technology have become so polarised that they block a fruitful dialogue aimed at finding room for action. On the one hand, a technocratic extreme pursues technological achievement without regard for its ethical or anthropological implications; on the other, a radical opposition to technology clearly overlooks the benefits it brings to areas that are now critical healthcare among them.

The constant tension between these two positions hinders ethical reflection, preventing a balanced appraisal of both the risks and the benefits that technological development offers. Along these lines, it is highly likely that the existing polarization is strongly amplified by today’s digital environment, where social networks and online platforms promote the flow of information not for its content or relevance, but rather for its popularity or its affinity with the user (Kubin & von Sikorski, 2021).

Considering the foregoing, it is essential to embrace technology as part of our reality while establishing ethical boundaries that safeguard human dignity and keep the person at the centre of our concerns. This would allow us to adopt an

ethical stance that assesses each technological development and intervention considering a proper understanding of the human person.

In this context of exponential technological growth, it is essential to engage in ethical reflection that acknowledges the importance of promoting and safeguarding human dignity. Moreover, we must recognise the complexity of caregiving, vulnerability as a constituent part of human nature, and intrinsic ontological dignity.

On the other hand, care is more than a technical action; it is a relational, dynamic, and deeply human experience. It involves recognition, identification, and, in many cases, mutual transformation between the caregiver and the care recipient. These aspects—essential to a full understanding of care—are difficult to replicate with a robot.

However, given the growing demand and the shortage of human caregivers, it is likely that in certain specific contexts the use of robots will prove inevitable and, at the same time, beneficial. In such cases, it will be essential to delineate clearly the appropriate scenarios for their deployment and the ethical boundaries of their actions.

Although technology may continue to improve and expand its functional capabilities, the ability to empathise with another person and share the experience of care remains

an irreplaceable challenge. This calls for deep ethical deliberation on the use of robots, taking into account not only their functions but also the characteristics and needs of the people they are intended to serve.

Therefore, it is essential to foster a fruitful dialogue that tackles these challenges from a person-centred perspective and affirms the intrinsic value of the human relationship in caregiving.

### **Human–Machine Interaction: Opportunities and Challenges in Healthcare**

Given the current limitations of robotics, we must ask whether the ethics of care—hitherto addressed by medical ethics and carried out by human physicians—can, to some extent, be replaced or at least complemented by robot doctors. There are, indeed, tangible advantages in using carebots to relieve human caregivers: they can accompany patients during lonely hospital admissions, dispense medication to different patients, and monitor their clinical parameters (Yew, 2021). In fact, studies have reported that some patients may prefer interacting with robots rather than human companions (Dautenhahn & Werry, 2004). Nevertheless, not every aspect of the large-scale inclusion of intelligent automata in healthcare—and in other sectors—can be resolved so readily.

One key issue is every individual's right to decide whether to use this technology, with decisions negotiated rather than imposed (Deusdad, 2024). Accordingly, a patient's consent to the use of carebots must be voluntary and unequivocal. Moreover, being assisted by intelligent automata under a physician's supervision is not the same as having the automaton act as the physician itself. To complicate matters further, we must consider what to do when people are mentally incapable of giving consent, such as very young children or elderly patients with advanced dementia (Ienca et al., 2016). Many argue that no one should be forced to be treated by an intelligent automaton, yet even this scenario seems unrealistic if we acknowledge the ongoing shrinkage of the healthcare workforce (Marco et al., 2021).

Sparrow and Sparrow (2006) have argued that robots are incapable of meeting the social and emotional needs of older adults under their care. We believe this is not always the case, yet developers and ethicists must explore different options to ensure that hospital care robots perform as effectively as possible. One major concern lies in the "art" of robot design: the way we build them affects the level of trust users who spend time with them will place in the device. Robots made of soft materials, small in size, and sporting zoomorphic features are generally better accepted by the average user (Coeckelbergh et al., 2016).

The question becomes even more complex when we ask whether we can truly trust and delegate to a robot. In the abstract:

trust is a decision by A to delegate to B some aspect of importance to A in achieving a goal, in which A's decisions are designed and implemented with the assumption that there is a high probability that B will behave as expected. (Yew, 2021, p. 634)

This has far-reaching implications. First, robots must be built with an ethical architecture and be rigorously tested to ensure they are as safe as possible. Only a robot that is perceived as “safe and robust” will be taken seriously by users (Salem et al., 2015).

At the same time, we must reflect on the fact that our conception of the intelligent automaton strongly conditions the degree of trust and the sense of safety that authorities and society attach to interacting with it. We realise that most automatons are still at an early stage of development; nonetheless, should a truly “intelligent automaton” emerge—one endowed with strong artificial intelligence, that is, with the capacities for learning, self-awareness, innovation, autonomy, and even sentience (Ruiz Osuna, 2025, p. 531)—our assumptions would likely need to be reframed under different paradigms.

The question of whether we can trust a robot can be answered in various ways depending on whether we regard the automaton as a “thing” or as a “someone” (Saerens, 2020). Indeed, if we acknowledge that robots may eventually become integrated into society, the prospect of placing trust in them and delegating tasks to them becomes more acceptable.

For instance, in Europe robots are predominantly conceived as mere instruments—nothing more than machines—whereas in traditional African thought personhood is fundamentally relational: what makes a person a person is his or her interconnectedness with others and with the community. In Japan, Shinto—the native religion—is a form of animism that attributes spirits, or *kami*, to all things: not only to human beings but also to animals, mountains, and everyday objects. Human beings are not above nature but part of it. The Shinto belief that “everything possesses spirits” refers to a powerful, impersonal force that is present in people and animals and can also be symbolised in objects, endowing them with power and sacredness that can be transmitted, inherited, or lost (Mori, 1981). From a Buddhist perspective, the fact that the automaton is a mechanical instrument does not separate it from the Buddha’s divine nature, because in Buddhism all things qualitatively contain the divine Being (Vallverdú, 2011, p. 261).

In the West, as noted earlier, the answer can be more complicated. What does seem clear is that users may feel



empathy or compassion toward robots depending on how advanced they are—or simply on whether the robot possesses a physical body (Ladak et al., 2024). In this context, so-called “AI ethics” usually focuses on how we build AI or how it ought to behave in its interactions with us; rarely does it ask whether we humans have obligations toward these entities, or whether they deserve moral consideration (Ruiz Osuna, 2019). Even so, some scholars argue that, as robots become more sophisticated, they should be granted certain rights—such as protection from physical harm (Mamak, 2022)—and that specific crimes against robots ought to be established (Mamak, 2023).

Whether we regard robots as moral agents or merely as “things,” there are pressing issues that must be addressed to achieve proper human–machine interaction. One of these concerns is the limits of formal verification. This limitation is echoed in recent work by Dickson (2024), who surveys the infeasibility of formally verifying modern, opaque AI models and argues that safety guarantees are often attainable only within simplified simulation environments. Accordingly, Seshia et al. (2022) advocated layered assurance architectures because complete formal verification is infeasible for systems operating in dynamic, unstructured settings. Taken together, these points support Floridi’s (2025) observation that “formal correctness scales only with strict constraints on environmental complexity”.

However, this is not the only limitation we must face. AI systems must be highly specific when performing caregiving tasks. In this regard, it is not enough that large language models (LLMs) have passed the Turing test (Jones & Bergen, 2025); even today, their way of “thinking” is not always satisfactory enough to entrust them with tasks without human oversight (Stapelfeldt, 2021). Moreover, whereas the configurations available to a machine are finite, the soul—Carabantes López (2016) reminds us—is a universal instrument capable of infinite combinations.

We must also address the question of exactly what kind of ethics we want robots to embody. Behaviour that is entirely for the benefit of others is sometimes referred to as “true altruism” (Daly et al., 2019). However, humans are rarely completely altruistic, which raises the issue of whether an intelligent automaton should be designed according to an “ideal” model of the average human or, instead, should realistically grasp the intricacies of human behavior—for example, understanding why people lie and when lying is socially acceptable (Serota et al., 2022). Indeed, as provocative as it may sound, the robot itself might need to know in which contexts a lie is permissible and, when necessary, lie (Redmalm et al., 2024).

An example of this can be found in the medical context of conveying diagnostic test results to patients. In most cases, patients feel more comfortable with robotic interactions in

that setting. However, when the news is not encouraging, patients generally prefer a human rather than an intelligent automaton to deliver the information (Hoorn & Winter, 2018). The perceived lack of empathy in robotic communication likely causes a degree of unease, as does the robot's limited ability to use the "white lies" that might soften the patient's stress when the results are unfavorable.

However, we have never faced a scenario of prolonged human-machine interaction; therefore, the long-term consequences of living alongside intelligent automata—particularly when they are highly autonomous and advanced—remain unclear. However, the first effects of such coexistence are already emerging: it is commonly found that children suffer distress and grief when separated from their robot companions (Yew, 2021). Nor do we know whether granting intelligent automata complete freedom of speech with humans would be dangerous, although Isaac and Bridewell (2017) argue that deception-capable robots can be ethical even when telling outright lies, and that the ethicality of human or robot communication must be assessed in light of its underlying motives, such as the pursuit of pro-social goals and functions.

However, the issue becomes more complex when dealing with vulnerable individuals—children, people with dementia, or patients with mental disorders—who may have difficulty discerning certain aspects of reality. In such cases,

the risk of disappointment emerges: both users and the robot's programmers may hold overly high expectations about the outcome of human-machine interaction. As Yew (2021) cautions,

I would argue that deception by care bots with regard to vulnerable persons may be allowed only if (a) the intention of the developer is ethical with respect to the care recipients, (b) the consequences are indeed positive for the care recipient, (c) there are no viable alternatives to the use of deception; and (d) the extent of infringement of the care recipient's autonomy is not more than that via other means. The additional two requirements (c) and (d) would further protect the vulnerable children and elderly from deceptive robots. (p. 632)

At least two patient scenarios must be considered. First are those with intact judgment—for instance, mental-health patients who still recognize that robots are not persons, even if we attribute some form of moral agency to them (Ruiz Osuna, 2025). Second, and more delicate, are cases in which the patient lacks full mental capacity to consent. Is it dangerous for a child to treat a robot with affection and grieve when it breaks down or is switched off? Is it ethical for users to form emotional bonds with intelligent automatons? Is it problematic when a person with dementia assumes that a robot is alive?

All these questions are far from simple because there is no consensus on what a robot is from an ontological standpoint. If, as Parks (2010) argues, carebots lack human attributes such as compassion, then—as Coeckelbergh et al. (2016) suggest—we should design and use robots according to the principle of “supervised autonomy.” Conversely, if we believe that an intelligent automaton meeting certain criteria (Ruiz Osuna, 2025) can be a moral agent, then users ought to treat it with respect, and trusting or even becoming attached to it would not be objectionable—much as we do with our pets.

At the present stage of scientific development, it is necessary to establish a course of action grounded in prudence, as many relevant scenarios lack significant empirical precedents. The long-term effects of cooperation with, or even delegation to, assistive robots in the medical field are unknown, given that their role has been largely marginal (Sparrow & Sparrow, 2006).

Advances in artificial intelligence and embodied artificial intelligence have enabled—and will increasingly enable—the broader implementation of these technologies in healthcare systems. However, given the many unresolved questions that remain, it does not seem appropriate to delegate to, or fully trust, systems that we do not yet completely understand (Seshia et al., 2022). Any implementation should therefore proceed gradually, under strict psychological and medical supervision, with tasks

assigned to robots remaining subject to meaningful human oversight (Yew, 2021).

## **Rethinking the Concept of Trust in the Age of AI: Can We Trust Robots?**

Trust is one of the most complex and fundamental phenomena in human experience, especially when projected into the field of robot-mediated care. Building on Innerarity's (2020) reflections on risk society, trust cannot be understood merely as an individual psychological attitude; rather, it is a social mechanism that enables us to navigate the uncertainty inherent in asymmetric power relations. In the context of caring for older adults with robots, this trust takes on particular dimensions: it entails not only confidence in the device's technical competence but also faith in the social, economic, and political systems that underpin its development and deployment (Innerarity, 2020).

From a broader philosophical standpoint, trust emerges as a fundamentally paradoxical act that reveals the human condition at its most vulnerable and potent (Carbajal, 2025a). Trust always involves a leap into the unknown, a temporary suspension of doubt that, as Simmel (1906) observed in his sociology of social forms, constitutes one of the most fundamental elements of social life. In the specific case of trusting care robots, this philosophical dimension is intensified because it confronts us with the question of

reciprocity: can genuine trust exist in a one-way relationship in which only one of the agents possesses moral consciousness? This query calls back to the phenomenological tradition which, from Husserl to Levinas (1961), has shown that authentic trust is grounded in recognizing the other as an alter ego, capable of responding ethically to our vulnerability.

Trust in care robots rests on a fundamental paradox that Harari (2018) explores at length in his analysis of the technological revolution: while these systems promise to ease the demographic care crisis, they simultaneously confront us with the question of what it means to be human in the age of artificial intelligence. In this context, trust is not merely an emotional response to a robot's effectiveness; it is a complex negotiation between our expectations of human caregiving and the algorithmic capacities of machines (Harari, 2018). This tension reveals that trust in robot-mediated care is intrinsically tied to our deepest conceptions of dignity, vulnerability, and reciprocity in caring relationships with other people.

Han (2017), in his critique of performance society, provides a crucial perspective for understanding how trust in robot-mediated care can become a subtle form of alienation. When we place our trust in automated care systems, we risk reducing care to a series of optimizable technical tasks, thereby losing sight of its existential and relational dimensions (Han, 2017). Paradoxically, trust in care robots

can erode our ability to trust ourselves and other human beings, turning care from an act of mutual recognition into an efficient yet depersonalized transaction.

Hence, implementing care robots is not merely a technical or economic choice but, above all, a political one, in the deep sense that Innerarity (2011) assigns to politics: the collective management of uncertainty and the construction of possible futures (Ruiz Osuna & Carbajal, 2025a). When a society decides to invest in care robots for its aging population (Innerarity, 2011). This decision involves a redistribution of resources and the emotional, physical, and social burdens of care, shifting them from the family and community sphere to the technological and commercial realm.

Consequently, from this perspective, trust in robot-mediated care becomes a tool of political legitimization that allows governments and societies to sidestep deeper debates about the social organization of care. As Harari (2016) argues in his analysis of algorithms and power, technological promises can serve to depoliticise fundamentally political problems: instead of asking why societies fail to build sustainable structures of human caregiving, trust is placed in technological solutions that appear neutral but actually embody specific biases and values about what it means to care and to be cared for. This “illusion of technological neutrality” enables collective responsibilities to be recast as seemingly technical decisions (Harari, 2016).



The political dimension of trust in robot-mediated care is also evident in how trust is distributed across society. It is no accident that care robots are marketed chiefly as solutions for vulnerable populations—older adults, people with dementia, and low-income families—while elites continue to enjoy premium human-care services. This differential distribution shows that trust in robot care is not universal but stratified by social class, thereby reflecting and perpetuating existing inequalities under the guise of neutral technological innovations (Han, 2014). In line with Han’s (2014) critique of neoliberal society, trust thus becomes a mechanism that normalizes a two-tier system of access to care.

Therefore, analyzing trust in robot-mediated care exposes fundamental tensions in contemporary society that extend far beyond technical or economic considerations. Trust in these systems cannot be reduced to a simple appraisal of their effectiveness; rather, it is a complex phenomenon that mirrors our collective anxieties about aging, loneliness, technology, and social responsibility (Innerarity, 2020; Harari, 2018; Han, 2017). In light of these authors’ insights, we can see that this trust is shot through with dynamics of power, commodification, and alienation that demand urgent critical scrutiny.

However, acknowledging these risks should not lead us to reject robotic technology in caregiving. Drawing on Innerarity’s (2011) notion of intelligently managing

uncertainty, we can cultivate reflective trust that harnesses the positive potential of such systems without succumbing to the alienation Han (2017) warns against. Care robots can serve as support tools that expand our human capacity to care, freeing time and energy for the more relational and emotional dimensions of caregiving, provided their deployment is guided by principles of complementarity rather than substitution. This approach requires a conscious political decision to use technology to strengthen, rather than erode, the fundamental human bonds at the heart of care (Ruiz Osuna & Carbajal, 2025b).

The fundamental question that emerges from this analysis is not whether we can trust care robots but what kind of society we are building when we choose to place that trust in automated systems and how we can do so in a way that preserves and strengthens (Innerarity, 2011; Han, 2014). Trust, understood as a collective political decision, confronts us with the opportunity to design hybrid care models that merge technological efficiency with human warmth. Robots would take on monitoring, physical assistance, and logistical tasks, while humans would focus on emotional support, active listening, and meaning-making. This transformative vision demands that trust in technology always remain subordinate to trust in our collective ability to build more caring and solidaristic societies (Carbajal, 2025b).

## Conclusions

Declining fertility rates and a steadily aging population, together with chronic shortages of both human and material resources in medicine, now pose serious challenges not only for governance but also for healthcare delivery itself.

Despite these difficulties, tomorrow's society must uphold the core principles of care ethics, grounded in patient autonomy and respect for human dignity. In this context, robotics and artificial intelligence are expected to play an increasingly prominent role in medicine.

However, as increasingly sophisticated medical robots enter the scene, doubts persist regarding the extent to which they can assume tasks that, historically, only human agents have been able to perform. Beyond their current technical limitations, many of which may eventually be overcome, the large-scale use of AI and robotics in hospitals raises thorny ethical and philosophical questions for which only scant empirical evidence is available.

At present, we do not know whether people can genuinely trust a robot and under what conditions. When the device is relatively simple, the answer seems straightforward; but matters grow murkier in the presence of a truly "intelligent automaton." At that point, any response is far

from universal because it hinges on what the individual user takes the robot to be in the first place.

Adding to the complexity, several studies have reported that patients often become attached to robots they encounter during therapy or hospital stays. Such situations could multiply exponentially if we delegate ever more medical tasks to advanced intelligent automatons. Here, consensus is lacking on a foundational issue: is it ethical for a robot to move beyond supporting clinicians and become a physician?

If we cannot guarantee a climate of trust with a robotic doctor, we may likewise be unable to guarantee an ethics of care mediated solely through an automaton, without human involvement. For the moment, robots may need to serve as increasingly capable assistants in hospitals, while a robust “machine ethics” framework is still taking shape—one that fosters genuine trust in human–machine interactions. Simultaneously, new studies are required in which patients spend extended periods with advanced robots so that both the intended and unintended effects of such prolonged contact can be rigorously evaluated.

## References

Alonso Galbán, P., Sansó Soberats, F. J., Díaz-Canel Navarro, A. M., Carrasco García, M., & Oliva, T. (2007). Envejecimiento poblacional y fragilidad en el adulto mayor. *Revista Cubana de Salud Pública*, 33(1).

- Aymerich-Franch, L., & Ferrer, I. (2021). Socially assistive robots' deployment in healthcare settings: A global perspective. *arXiv*. <https://arxiv.org/abs/2110.07404>
- Bartoli Presas, M. (2020). Persona y libertad: ¿Somos realmente libres? In A. V. Durruty Corral (Ed.), *El camino de la persona. Libro I: Antropología filosófica* (pp. 92–103). Ediciones Universidad San Sebastián.
- Bevilacqua, M. G. (2016). Caring about ethics of care: A new dimension. *Phenomenology and Mind*, (11), 114–119. [https://doi.org/10.13128/Phe\\_Mi-20111](https://doi.org/10.13128/Phe_Mi-20111)
- Bricker, D., & Ibbitson, J. (2019). *Empty planet: The shock of global population decline*. Crown.
- Busquets, E. (2019). *Ética del cuidado en ciencias de la salud*. Herder.
- Carabantes López, M. (2016). El pronóstico de Descartes sobre los problemas de la inteligencia artificial. *SCIO. Revista de Filosofía*, 12, 201–228. <https://dialnet.unirioja.es/servlet/articulo?codigo=5867803>
- Carbajal, I. (2025a). Epistemología sin sujeto: desubjetivación y saber en la era de la inteligencia artificial. *Estudios de Epistemología*, 22(1), 52–61. <https://estudiosdeepistemologia.ct.unt.edu.ar/article/view/491>
- Carbajal, I. (2025b). Explorando las fronteras digitales de la religión: Un análisis de las fronteras de la Inteligencia

- Artificial y la Teología. *Cuadernos de Teología*, 17, e6558. <https://doi.org/10.22199/issn.0719-8175-6558>
- Coeckelbergh, M., Pop, C., Simut, R., Peca, A., Stahl, B. C., Holvoet, J., & Vandemeulebroucke, T. (2016). A survey of expectations about the role of robots in robot-assisted therapy for children with ASD. *Science and Engineering Ethics*, 22(1), 47–65. <https://doi.org/10.1007/s11948-015-9649-x>
- Daly, J. E., Bremner, P., & Leonards, U. (2019). Robots in need: Acquiring assistance with emotion. In *Companion of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 706–708). IEEE. <https://doi.org/10.1109/HRI.2019.8673081>
- Dautenhahn, K., & Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1), 1–35.
- Deusdad, B. (2024). Ethical implications in using robots among older adults living with dementia. *Frontiers in Psychiatry*, 15, Article 1436273. <https://doi.org/10.3389/fpsy.2024.1436273>
- Dickson, A. (2024). Limitations on formal verification for AI safety. *AI Alignment Forum*. <https://www.alignmentforum.org/posts/B2bg677TaS4cmDPzL/>
- Floridi, L. (2025). A conjecture on a fundamental trade-off between certainty and scope in symbolic and

- generative AI. *Philosophy & Technology*, 38(3), 93.  
<https://doi.org/10.1007/s13347-025-00927-z>
- Hajizadeh, A., Hafezi, R., Torabi, F., Akbari Sari, A., & Tajvar, M. (2025). Consequences of population ageing on health systems: A conceptual framework for policy and practice. *Ethiopian Journal of Health Sciences*, 35(1), 51–62. <https://doi.org/10.4314/ejhs.v35i1.8>
- Han, B.-C. (2014). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder* (A. Bergés, Trans.). Herder.
- Han, B.-C. (2017). *La sociedad del cansancio* (2nd expanded ed.; A. Saratxaga Arregi & A. Ciria, Trans.). Herder.
- Harari, Y. N. (2016). *Homo Deus: Breve historia del mañana* (J. Ros, Trans.). Debate.
- Harari, Y. N. (2018). *21 lecciones para el siglo XXI* (J. Ros, Trans.). Debate.
- Hoorn, J. F., & Winter, S. D. (2018). Here comes the bad news: Doctor robot taking over. *International Journal of Social Robotics*, 10(4), 519–535.  
<https://doi.org/10.1007/s12369-017-0455-2>
- Ienca, M., Jotterand, F., Vica, C., & Elger, B. (2016). Social and assistive robotics in dementia care: Ethical recommendations for research and practice. *International Journal of Social Robotics*, 8, 565–573.
- IndexMundi. (2020, December 1). *Mundo – Tasa de mortalidad (muertes/1 000 habitantes) – Cuadros de datos históricos anuales*.

[https://www.indexmundi.com/g/g.aspx?v=26&c=xx  
&l=es](https://www.indexmundi.com/g/g.aspx?v=26&c=xx&l=es)

- INE. (2025). *Estadística de nacimientos: Movimiento Natural de la Población (datos definitivos de 2023)*. Instituto Nacional de Estadística. [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736177007&menu=ultiDatos&idp=1254735573002](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177007&menu=ultiDatos&idp=1254735573002)
- Innerarity, D. (2011). *La democracia del conocimiento: Por una sociedad inteligente*. Paidós.
- Innerarity, D. (2020). *Pandemocracia: Una filosofía de la crisis del coronavirus*. Galaxia Gutenberg.
- Isaac, A. M. C., & Bridewell, W. (2017). White lies on silver tongues: Why robots need to deceive (and how). In P. Lin, K. Abney, & R. Jenkins (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 103–118). Oxford University Press.
- Ismail, Z., Wan Ahmad, W. I., Hanin Hamjah, S., & Astina, I. K. (2021). The impact of population ageing: A review. *Iranian Journal of Public Health*, 50(12), 2451–2460. <https://doi.org/10.18502/ijph.v50i12.7927>
- Jones, C. R., & Bergen, B. K. (2025). Large language models pass the Turing test (arXiv Preprint No. 2503.23674). *arXiv*. <https://arxiv.org/abs/2503.23674>
- Kallestrup-Lamb, M., Marin, A. O. K., Menon, S., & Søgaaard, J. (2024). Aging populations and expenditures on health. *The Journal of the Economics of Ageing*, 29,



<https://doi.org/10.1016/j.jeo.2024.100518>

- Kant, I. (2002). *Fundamentación para una metafísica de las costumbres* (M. García Morente, Trans.). Alianza Editorial. (Original work published 1785)
- Kittay, E. F., Jennings, B., & Wasunna, A. A. (2005). Dependency, difference and the global ethic of long-term care. *The Journal of Political Philosophy*, 13(4), 443–469. <https://doi.org/10.1111/j.1467-9760.2005.00231.x>
- Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), 188–206. <https://doi.org/10.1080/23808985.2021.1976070>
- Ladak, A., Harris, J., & Anthis, J. R. (2024). Which artificial intelligences do people care about most? A conjoint experiment on moral consideration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Article 287, pp. 1–11). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642403>
- Leocadie, M.-C., Morvillers, J.-M., Pautex, S., & Rothan-Tondeur, M. (2020). Characteristics of the skills of caregivers of people with dementia: Observational study. *BMC Family Practice*, 21, Article 149. <https://doi.org/10.1186/s12875-020-01218-6>

- Levinas, E. (1961). *Totalité et infini: Essai sur l'extériorité*. Martinus Nijhoff.
- Mamak, K. (2022). Should violence against robots be banned? *International Journal of Social Robotics*, 14(4), 1057–1066. <https://doi.org/10.1007/s12369-021-00852-z>
- Mamak, K. (2023). *Robotics, AI and criminal law: Crimes against robots*. Routledge. <https://doi.org/10.4324/9781003331100>
- Marco, C. A., Courtney, D. M., Ling, L. J., Salsberg, E., Reisdorff, E. J., Gallahue, F. E., Suter, R. E., Muelleman, R., Chappell, B., Dowling Evans, D., Vafaie, N., & Richwine, C. (2021). The emergency medicine physician workforce: Projections for 2030. *Annals of Emergency Medicine*, 78(6), 726–737. <https://doi.org/10.1016/j.annemergmed.2021.05.029>
- Maresova, P., Javanmardi, E., Barakovic, S., Barakovic Husic, J., Tomsone, S., Krejcar, O., & Kuca, K. (2019). Consequences of chronic diseases and other limitations associated with old age – A scoping review. *BMC Public Health*, 19, Article 1431. <https://doi.org/10.1186/s12889-019-7762-5>
- Molas-Tuneu, M., Jerez-Roig, J., Minobes-Molina, E., Coll-Planas, L., Escribà-Salvans, A., Farrés-Godayol, P., Yildirim, M., Rierola-Fochs, S., Goutan-Roura, E., Leandro Bezerra de Souza, D., Skelton, D. A., & Naudó-Molist, J. (2023). Social and emotional loneliness among older people living in nursing homes in Spain:

- A cross-sectional study. *Anales de Psicología / Annals of Psychology*, 39(3), 465–477. <https://doi.org/10.6018/analesps.530641>
- Mori, M. (1981). *The Buddha in the robot: A robot engineer's thoughts on science and religion*. Kōsei Publishing.
- Parks, J. A. (2010). Lifting the burden of women's care work: Should robots replace the "human touch"? *Hypatia*, 25(1), 100–120. <https://doi.org/10.1111/j.1527-2001.2009.01065.x>
- Pettersen, T. (2011). The ethics of care: Normative structures and empirical implications. *Health Care Analysis*, 19(1), 51–64. <https://doi.org/10.1007/s10728-010-0163-7>
- Postigo Solana, E. (2022). Naturaleza humana y transhumanismo. *Pensamiento. Revista de Investigación e Información Filosófica*, 78(298), 523–534. <https://doi.org/10.14422/pen.v78.i298.y2022.003>
- Redmalm, D., Iversen, C., & Persson, M. (2024). Can robots lie? A post-humanist approach to robotic animals and deceptive practices in dementia care. *Journal of Aging Studies*, 71, 101272. <https://doi.org/10.1016/j.jaging.2024.101272>
- Rodrigues Dixe, M. d. A. C., Teixeira, L. F. d. C., Areosa, T. J. T. C., Frontini, R. C., Peralta, T. d. J. A., & Querido, A. I. F. (2019). Needs and skills of informal caregivers to care for a dependent person: A cross-sectional study.

- Ruiz Osuna, P. (2019). *La personalidad y responsabilidad de los robots* [Doctoral dissertation, Universitat Rovira i Virgili].
- Ruiz Osuna, P. (2025). *La personalidad jurídica de los autómatas inteligentes*. Aranzadi La Ley.
- Ruiz Osuna, P., & Carbajal, I. (2025a). Nota de los editores. *Colloquia. Revista de Pensamiento y Cultura*, 12. <https://doi.org/10.31207/colloquia.v12i1.207>
- Ruiz Osuna, P., & Carbajal, I. (2025b). ¿Pueden salvarse los robots? Análisis interpretativo de la consideración moral de los autómatas a la luz de la teología. *Cuestiones Teológicas*, 53(119), 1–17. <https://doi.org/10.18566/cueteo.v53n119.a05>
- Saerens, P. (2020). Le droit des robots, un droit de l'homme en devenir? *Communication, Technologies et Développement*, 8, 1–17.
- Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot? Effects of error, task type and personality on human–robot cooperation and trust. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 141–148). ACM. <https://doi.org/10.1145/2696454.2696497>
- Serota, K. B., Levine, T. R., & Docan-Morgan, T. (2022). Unpacking variation in lie prevalence: Prolific liars,

- bad lie days, or both? *Communication Monographs*, 89(3), 307–331.  
<https://doi.org/10.1080/03637751.2021.1985153>
- Seshia, S. A., Sadigh, D., & Sastry, S. S. (2022). Toward verified artificial intelligence. *Communications of the ACM*, 65(7), 46–55. <https://doi.org/10.1145/3503914>
- Simmel, G. (1906). The sociology of secrecy and of secret societies. *American Journal of Sociology*, 11(4), 441–498. <https://doi.org/10.1086/211418>
- Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged-care. *Minds and Machines*, 16(2), 141–161. <https://doi.org/10.1007/s11023-006-9030-z>
- Stapelfeldt, R. (2021). Can there be a dumb super-intelligence? A critical look at Bostrom's notion of superintelligence. *Academia Letters*, Article 2076. <https://doi.org/10.20935/AL2076>
- Ten Have, H. (2016). *Vulnerability: Challenging bioethics*. Routledge.
- Vallverdú, J. (2011). Review: Masahiro Mori, The Buddha in the Robot. *Enrahonar: An International Journal of Theoretical and Practical Reason*, 47, 261–262. <https://doi.org/10.5565/rev/enrahonar.180>
- Van Nistelrooij, I., & Visse, M. (2019). Me? The invisible call of responsibility and its promise for care ethics: A phenomenological view. *Medicine, Health Care and Philosophy*, 22(2), 275–285. <https://doi.org/10.1007/s11019-018-9873-7>

- Varkey, B. (2021). Principles of clinical ethics and their application to practice. *Medical Principles and Practice*, 30(1), 17–28.  
<https://doi.org/10.1159/000509119>
- Yew, G. C. K. (2021). Trust in and ethical design of carebots: The case for ethics of care. *International Journal of Social Robotics*, 13(4), 629–645.  
<https://doi.org/10.1007/s12369-020-00653-w>

# Ética de la rotación laboral en la era de la Inteligencia Artificial: ¿condición para el crecimiento económico o forma de exclusión?

*Ethics of Labor Turnover in the Age of Artificial  
Intelligence: A Condition for Economic Growth  
or a Form of Exclusion?*

**Leonardo Manríquez-Méndez<sup>5</sup>**

## **Resumen**

La inteligencia artificial (IA) está transformando el mercado laboral, generando un debate ético sobre su rol como motor de crecimiento económico o como fuente de exclusión.

---

<sup>5</sup> Universidad Católica de la Santísima Concepción.  
<https://orcid.org/0009-0008-5028-5848>

Este estudio analiza su impacto en la rotación laboral, evidenciando una dualidad en sus efectos. Por un lado, la IA impulsa la productividad y la creación de empleos en sectores como ciberseguridad, salud digital y marketing. Por otro, desplaza trabajadores en distintos niveles de cualificación, afectando especialmente a grupos vulnerables y no-STEM, lo que plantea desafíos en términos de dignidad humana y justicia distributiva. Desde una perspectiva neutral, se sostiene que sus efectos dependen del rol de las instituciones, en particular en educación y regulación. Se concluye que la IA puede fomentar el crecimiento económico si se acompaña de políticas inclusivas, aunque persisten limitaciones asociadas a la incertidumbre empírica y la falta de evidencia longitudinal.

**Palabras clave:** Rotación laboral; crecimiento económico; exclusión tecnológica; justicia distributiva; inteligencia artificial.

## **Abstract**

Artificial intelligence (AI) is transforming the labor market, raising an ethical debate about its role as a driver of economic growth or a source of exclusion. This study analyzes its impact on labor turnover, highlighting a dual effect. On one hand, AI enhances productivity and creates jobs in sectors such as cybersecurity, digital health, and marketing. On the other, it displaces workers across different skill levels,



disproportionately affecting non-STEM and vulnerable groups, which raises concerns about human dignity and distributive justice. From a neutral perspective, its effects depend largely on the role of institutions, particularly in education and regulation. The findings suggest that AI can foster economic growth if accompanied by inclusive policies, although limitations remain due to empirical uncertainty and the lack of longitudinal evidence.

**Keywords:** Labor turnover; economic growth; technological exclusion; distributive justice; artificial intelligence.

## **Introducción**

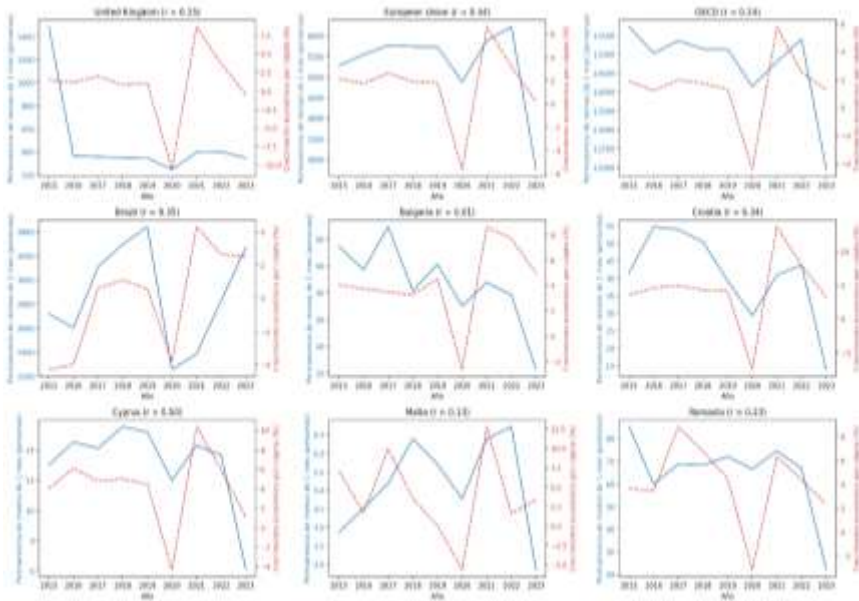
El dinamismo económico, caracterizado por un crecimiento acelerado, depende en gran medida de la rotación laboral, entendida como la entrada de nuevos trabajadores y empresas al mercado, así como la salida de aquellos menos competitivos. Esta relación positiva entre rotación y crecimiento económico se fundamenta en la capacidad de los mercados laborales para reasignar recursos hacia sectores más productivos (Siebert & Zubanov, 2009), fomentando la innovación y la adaptación a nuevas demandas (Coles & Mortensen, 2016). En la era de la Inteligencia Artificial (IA), estos mecanismos se intensifican, ya que la automatización acelera tanto la sustitución de trabajadores como la creación de nuevas oportunidades, transformando las estructuras laborales. En economías emergentes, donde la adopción de la

IA es creciente, esta relación plantea interrogantes tanto económicos como éticos.

La rotación laboral puede impulsar el crecimiento económico, como lo sugieren las correlaciones iniciales entre la permanencia en el empleo y el crecimiento económico per cápita en países de la OCDE, la Unión Europea, el Reino Unido y otras economías no pertenecientes a la OCDE (véase la Figura 1). Sin embargo, este fenómeno también genera exclusión, especialmente para los trabajadores desplazados por tecnologías avanzadas. Esto desafía la noción de que el trabajo dignifica a la persona humana y plantea dilemas éticos en torno a la justicia distributiva y el derecho al trabajo.

## Figura 1

*Permanencia laboral (menos de 1 mes) v.s. Crecimiento Económico per cápita (2015-2023)*



*Nota. r representa el coeficiente de correlación de Pearson entre la permanencia laboral y crecimiento económico per cápita para cada uno de los países o grupos de países.*

La noción de que el trabajo dignifica al ser humano es central para comprender los dilemas éticos de la exclusión laboral en la era de la IA. Según Sen (1999) el desarrollo económico debe priorizar la expansión de las capacidades

humanas, entendiendo el acceso a un trabajo digno como una fuente de autorrealización y de justicia distributiva. La exclusión laboral, como la inducida por la automatización, limita estos atributos, restringiendo no solo las oportunidades económicas, sino también el sentido de identidad y dignidad que el trabajo confiere. En una línea similar, Arendt (2013) descompone el concepto de trabajo como fuente de producción, identidad y dignidad humana, y advierte que las tendencias desreguladas de automatización pueden conducir a la deshumanización del trabajo.

En este sentido, Janzen y DeRoo (2022) exploran la posibilidad de que las máquinas realicen la totalidad del trabajo humano, lo que conduciría a la deshumanización del trabajo. Los autores concluyen que este escenario no es deseable, ya que ciertas tareas requieren elementos esencialmente humanos que las máquinas no pueden replicar adecuadamente, como la empatía o las conexiones interpersonales.

Aunque una deshumanización total del trabajo aún parece lejana, estas nociones cobran relevancia ante resultados como los de Frey y Osborne (2017) quienes estiman que cerca del 47% de los empleos en economías desarrolladas están en riesgo de automatización, incluyendo tareas susceptibles de ser reemplazadas por IA. Por su parte, Harari (2017) sostiene que la IA y la automatización pueden desplazar a millones de trabajadores, creando una clase *inútil*:

un grupo de personas cuyos empleos son reemplazados y que no poseen habilidades relevantes para los nuevos mercados laborales. En este contexto, la rotación laboral acelerada por la IA, si no se acompaña de políticas que garanticen la reinserción laboral, amenaza los principios de equidad y dignidad, exigiendo una reflexión ética sobre sus implicaciones.

A pesar de las preocupaciones y predicciones pesimistas sobre la relación entre automatización y pérdida de empleos, Acemoglu et al. (2023) realiza un análisis multisectorial en el que destaca que las nuevas tecnologías transformarán el mercado laboral de manera gradual y manejable. Por ejemplo, en el contexto de la Industria 4.0, los avances en digitalización, como la robótica colaborativa y el internet de las cosas, han mostrado progresos significativos; sin embargo, su adopción sigue siendo más lenta en las pequeñas y medianas empresas. Asimismo, en el ámbito de la logística, la automatización —como el uso de robots móviles— se enfoca en asistir a los trabajadores, no en reemplazarlos.

En este sentido, y siguiendo a Dwivedi et al. (2021) resulta pertinente referirse al concepto de cambio en los roles laborales, ya que, mediante la automatización de tareas repetitivas y basadas en reglas, los trabajadores podrían enfocarse en funciones más creativas y orientadas al pensamiento crítico. Esto les permitiría ascender en la cadena de valor, fortaleciendo la relación triádica entre cliente,

tecnología y empleado (Vomberg et al., 2024). En medio de este debate, coexisten posturas contrastantes respecto al impacto de la inteligencia artificial en el mercado laboral: mientras algunos subrayan sus beneficios en términos de productividad y generación de empleo, otros advierten sobre su capacidad para provocar desempleo y aumentar las desigualdades. A estas visiones se suma una postura más neutral o matizada, que reconoce tanto las oportunidades como los riesgos, subrayando la necesidad de análisis contextuales y basados en evidencia. Estas perspectivas divergentes no solo enriquecen el debate, sino que también plantean interrogantes clave, como si la IA debe considerarse una vía hacia el crecimiento económico o, por el contrario, un factor de exclusión laboral. Esta es, precisamente, la pregunta central que orienta el presente trabajo.

Este estudio busca comprender si los efectos de la inteligencia artificial sobre el mercado laboral son predominantemente favorables o desfavorables, en particular en lo que respecta a su papel como motor de crecimiento económico o como factor de exclusión laboral. Para ello, se toma como punto de partida la evidencia empírica que muestra una relación entre rotación laboral y crecimiento económico, lo que permite plantear la hipótesis de que la automatización impulsada por la IA podría estar influyendo en dicha dinámica. A partir de esta premisa, el objetivo principal es analizar en qué medida los impactos laborales de la IA constituyen una condición para el desarrollo o una forma

de exclusión estructural en los mercados laborales contemporáneos.

La metodología empleada consiste en una revisión crítica de los principales estudios que respaldan, en primer lugar, una postura favorable respecto al impacto de la IA en el empleo; en segundo lugar, aquellos que advierten sobre sus efectos negativos; y, finalmente, se incorpora una perspectiva más neutral y contextual, basada en los aportes recientes de Daron Acemoglu y Simon Johnson, quienes proponen una visión intermedia que enfatiza el papel de las decisiones políticas y del diseño institucional en la configuración de dichos impactos.

En cuanto a la estructura del trabajo: en la sección 2 se presentan los materiales y métodos; en la sección 3, los resultados; y en la sección 4, las conclusiones en torno al debate sobre el equilibrio entre el progreso económico y las consideraciones éticas.

## **Materiales y métodos**

Para explorar la compleja relación entre la IA y el mercado laboral, con un enfoque particular en la rotación laboral, este artículo realiza un análisis comparativo de artículos académicos relevantes que buscan responder a la pregunta: ¿cómo afecta la IA al mercado laboral? Los artículos seleccionados presentan perspectivas diversas sobre el

impacto de la automatización impulsada por la IA, abarcando desde evidencia empírica sobre los efectos en el empleo hasta discusiones teóricas sobre vulnerabilidades sectoriales y demográficas. El análisis se estructura en tres subsecciones: la primera examina los aspectos de exclusión laboral derivados de la adopción de la IA, centrándose en la sustitución de empleos y sus impactos negativos; la segunda evalúa los aspectos positivos de la IA en el mercado laboral, como la creación de nuevos roles y el aumento de la productividad; y la tercera ofrece una visión neutral, poniendo el foco en la institucionalidad como eje articulador del debate. Mediante una comparación cualitativa, esta investigación sintetiza las preguntas centrales, las respuestas y las contribuciones de cada artículo, identificando puntos de consenso y divergencia.

### ***Desplazamiento laboral por automatización con IA***

La adopción de la IA plantea preocupaciones significativas sobre el desplazamiento y sustitución de la mano de obra. Estudios recientes han abordado estas dinámicas, destacando tanto los efectos de sustitución laboral como las oportunidades de creación de nuevas tareas, por ejemplo, Demirci et al. (2025) cuantifica el impacto de la IA generativa en plataformas de *freelancing*, evidenciando una disminución del 21% en las ofertas de empleo para trabajos de escritura y programación tras la introducción de ChatGPT, y un 17% en trabajos relacionados con la creación de imágenes



tras la aparición de IA generadora de imágenes, este análisis subraya la vulnerabilidad de tareas creativas y técnicas frente a la automatización avanzada. Por su parte, Huang (2024) explora el impacto de la IA en Estados Unidos, encontrando que las zonas con mayor exposición a la IA experimentan una reducción en la tasa de empleo-población<sup>6</sup>, afectando especialmente a sectores como la manufactura y los servicios de baja cualificación, así como a trabajadores de habilidad media, no STEM<sup>7</sup>, hombres y grupos etarios extremos (16-25 y mayores de 46 años), sus hallazgos sugieren la necesidad de políticas de recualificación para mitigar una sustitución laboral heterogénea.

En un análisis histórico-comparativo, Deming et al. (2024) destacan que la disrupción tecnológica actual, impulsada por la IA, difiere de períodos previos al romper con la polarización laboral. Observan una disminución en empleos de salarios bajos y medios, un estancamiento en el crecimiento de empleos de servicios de baja remuneración, un aumento del 50% en empleos STEM desde 2010 (especialmente en ocupaciones de *software* e informática) y una caída del 25% en el empleo en ventas minoristas, atribuida a avances en el comercio en línea, estas tendencias

---

<sup>6</sup> Esta razón puede ser considerada como un símil de la tasa de ocupación laboral en una economía.

<sup>7</sup> STEM viene desde la abreviación de *Science, Technology, Engineering y Mathematics*.

reflejan el impacto temprano de la IA en ocupaciones administrativas y de ventas. Sin embargo, los autores advierten que la IA debe ser visto como una tecnología de propósito general (GPT por su traducción al inglés), como la energía a vapor y la electricidad, lo que implica que su impacto se producirá a través de largas décadas, por lo tanto, es demasiado pronto para evaluar su impacto total. En el contexto chino, Wang et al. (2024) muestran que la IA optimiza la estructura industrial, pero reduce los empleos de baja cualificación en un 0.28% por cada 1% de aumento en la adopción de IA, mientras que los empleos de alta cualificación crecen, evidenciando una polarización en el mercado laboral.

Suleyman y Bhaskar (2023) sostienen que los modelos avanzados de IA superarán a los humanos en tareas cognitivas como redacción, traducción y servicio al cliente, volviendo obsoletos muchos empleos de cuello blanco a corto plazo. Aunque surgirán nuevas oportunidades laborales, estas no compensarán la magnitud ni la rapidez de las pérdidas, y las fricciones del mercado laboral (habilidades, geografía e identidad) complicarán la transición de los trabajadores. Además, la automatización reducirá los ingresos fiscales, afectando servicios públicos y programas de bienestar, y aumentará la desigualdad, la inseguridad laboral y el descontento social, especialmente en regiones con reemplazo lento de empleos.

A nivel empresarial, Acemoglu et al. (2020) analizan la adopción de robots industriales en Francia, encontrando que las empresas que implementan robots incrementan su productividad en un 20%, pero reducen el empleo en ocupaciones no cualificadas en un 10%, sin cambios en el empleo total. Finalmente, Acemoglu y Restrepo (2018) ofrecen un marco teórico basado en tareas para comprender los efectos de la automatización y la IA, destacando un efecto de desplazamiento que reduce la demanda de trabajo y los salarios. Sin embargo, la creación de nuevas tareas intensivas en mano de obra puede contrarrestar este impacto, aunque enfrenta limitaciones como la falta de habilidades para nuevas tecnologías y una posible sobre adopción de automatización en detrimento de tecnologías que potencian la productividad. En conjunto, estos estudios evidencian que, si bien la IA impulsa la productividad y la creación de empleos de alta cualificación, genera desafíos significativos para los trabajadores menos cualificados y requiere estrategias de adaptación para equilibrar sus impactos en el mercado laboral.

### ***Beneficios laborales de la automatización con IA***

En contraste a la sustitución laboral, coexiste una postura de que la IA está transformando el mercado laboral al generar nuevas oportunidades, mejorar la productividad y fomentar la creación de roles innovadores y contrarrestando los temores del desplazamiento masivo. Machovec et al.

(2025) destacan que la IA impulsa la demanda de roles que requieren habilidades complementarias, como el análisis de riesgos en finanzas, proponiendo un marco para integrar estos impactos en proyecciones laborales y enfatizando en la colaboración humano-máquina en sectores como computación, ingeniería y leyes. De manera similar, Hampole et al. (2025) compilan investigaciones de la conferencia *National Bureau of Economic Research 2025* y muestran que la IA fomenta empleos en ciberseguridad, salud y marketing digitales, generando un efecto de reinstalación que reubica a los trabajadores en roles creativos y estratégicos, junto con la necesidad de formación en habilidades digitales. En países en desarrollo, Shen y Zhang (2024) encuentran que la adopción de robots industriales en China incrementa el empleo al mejorar la productividad y fomentar la aglomeración virtual para la colaboración remota, contrarrestando el desplazamiento laboral a través de un estudio empírico en 30 provincias.

Por otro lado, Cazzaniga (2024) señala que los trabajadores con educación universitaria están mejor posicionados para aprovechar la IA generativa, lo que reduce barreras para jóvenes educados y permite complementar roles cognitivos. Proponen políticas de capacitación y redes de seguridad social para una transición inclusiva. Acemoglu et al. (2023) argumentan que la IA no provocará un desplazamiento masivo, sino una transformación gradual de los empleos, ya que automatiza tareas específicas en lugar de

ocupaciones completas, permitiendo a los trabajadores enfocarse en actividades de mayor valor. Subrayan que este proceso, lento y extendido por décadas, ofrece tiempo para preparar a los trabajadores y ajustar las instituciones.

En roles de baja y media cualificación, un estudio de Brynjolfsson et al. (2023) demuestra que un asistente de IA generativa mejora la productividad en soporte al cliente en un 14% (hasta 34% para trabajadores menos cualificados), reduciendo la desigualdad salarial al mejorar el desempeño de estos grupos sin desplazar empleos. Finalmente, Song et al. (2023) evidencian que la IA generativa incrementa la productividad en el desarrollo de software en un 5.5% y la participación en un 5.4%, aunque con un aumento del 41.6% en el tiempo de coordinación, democratizando el acceso a roles técnicos mediante la integración de herramientas de IA en flujos de trabajo. Estos hallazgos sugieren que la IA, lejos de ser una amenaza, puede complementar el trabajo humano, promover la inclusión y generar un mercado laboral más dinámico e innovador con las políticas adecuadas.

### ***Más allá del optimismo y pesimismo: una visión institucional***

La IA está reconfigurando el mercado laboral al introducir cambios en la organización del trabajo y las demandas de habilidades. Desde una perspectiva neutral, los efectos de la IA no son inherentemente positivos ni negativos,

sino que están determinados por las decisiones de los actores sociales, económicos y políticos que guían su desarrollo e implementación (Acemoglu & Johnson, 2023). Esta tecnología tiene la capacidad de transformar tanto las oportunidades como los desafíos en el ámbito laboral, dependiendo de cómo se gestionen estas transformaciones.

La IA permite automatizar tareas rutinarias y analíticas, lo que puede incrementar la eficiencia en sectores como la industria, los servicios y la logística (Acemoglu & Jognson, 2024). Esta automatización puede liberar a los trabajadores de actividades repetitivas, permitiéndoles enfocarse en roles que demanden creatividad o habilidades interpersonales. Este proceso de transición refleja una reestructuración del mercado laboral que no es intrínsecamente favorable ni perjudicial.

Además, de acuerdo con Acemoglu et al. (2023) la IA impulsa la creación de nuevas oportunidades laborales en áreas como el desarrollo tecnológico, la gestión de datos y ciberseguridad. Este potencial para generar nuevos roles subraya la importancia de diseñar tecnologías que complementen las capacidades humanas, como se propone en enfoques que priorizan una IA al servicio de los trabajadores. Al mismo tiempo, la IA plantea retos para los sistemas de educación y formación, y la evidencia muestra que el impacto de las innovaciones tecnológicas depende de la capacidad de

las instituciones para fomentar la creación de nuevas tareas y facilitar la transición laboral (Acemoglu & Johnson, 2024).

El impacto de la IA en el mercado laboral no está predeterminado. Su trayectoria dependerá de cómo las sociedades equilibren las oportunidades de innovación con los desafíos de adaptación, a través de políticas que promuevan una distribución equitativa de los beneficios tecnológicos (Acemoglu & Johnson, 2023). Este proceso ya lo advertía Acemoglu y Robinson (2013) en su análisis sobre el desarrollo económico, donde concluye que la institucionalidad juega un rol fundamental en la adopción de tecnologías en contextos de cambio. Las sociedades que logren gestionar eficazmente sus instituciones serán las que puedan liderar los beneficios de la IA, garantizando un mercado laboral resiliente y sostenible.

## **Resultados**

El análisis comparativo de los artículos revisados en los dos extremos (impactos negativos y positivos) revela tendencias claras en la relación entre la IA y el mercado laboral, organizadas en dos categorías: los riesgos asociados con la sustitución laboral y las oportunidades derivadas de la creación de nuevos roles y el aumento de la productividad (Cuadro 1) Estas tendencias, analizadas en el contexto de la rotación laboral, se entrelazan con consideraciones éticas fundamentales, como la dignidad humana, la justicia

distributiva y el derecho al trabajo (Sen, 1999) lo que exige una reflexión sobre el equilibrio entre progreso económico y equidad social.

**Tabla 1**

*Tendencias identificadas entre IA y Mercado Laboral*

| <b>Riesgos y sustitución<br/>laboral</b>                                                                                                                                                                                                                  | <b>Creación de nuevos roles y<br/>aumento de productividad</b>                                                                                                                                                                                 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sustitución de empleos en múltiples niveles de cualificación: varios estudios muestran que la IA y los robots reducen empleos de baja cualificación (tanto en servicios como en manufactura) y podría convertirse en una amenaza para trabajos cognitivos | Creación de nuevos empleos y habilidades: la IA genera nuevos roles en áreas como ciberseguridad, salud digital, marketing y análisis de riesgos, aumentando la demanda de habilidades digitales y estratégicas.                               |
| Impacto desigual por demografía, geografía y sectores: el efecto de la IA varía según edad, género, habilidades y ubicación. Jóvenes, hombres y trabajadores no-STEM son                                                                                  | Complementariedad humano-IA y mejora de productividad: la IA no reemplaza ocupaciones completas, sino que automatiza tareas específicas, permitiendo a los trabajadores enfocarse en actividades de mayor valor y aumentando su productividad, |



**Riesgos y sustitución  
laboral****Creación de nuevos roles y  
aumento de productividad**

---

más afectados en áreas de alta exposición a IA.

especialmente en roles de baja y media cualificación.

Desplazamiento laboral, nuevas demandas laborales y fricciones: hay una caída en trabajos tradicionales (ventas minoristas y servicios) y un crecimiento en sectores STEM y de alta cualificación, impulsado por la adopción tecnológica. Esto generó fricciones sobre salarios.

Reducción de desigualdades y transición inclusiva: la IA puede ayudar a reducir brechas salariales y de oportunidades, favoreciendo especialmente a trabajadores jóvenes, menos cualificados o con educación universitaria, si se acompañan con políticas de recapacitación y apoyo social.

---

Nota. Elaboración propia.

Las tendencias presentadas revelan una dualidad en los efectos de la IA: mientras impulsa el crecimiento económico mediante la rotación laboral y la creación de nuevos roles, también genera exclusión para trabajadores menos cualificados y grupos demográficos vulnerables. Las consideraciones éticas, como la dignidad humana y la justicia distributiva, son centrales para abordar estos desafíos.

Las experiencias de países de la OCDE, la Unión Europea, el Reino Unido y países no pertenecientes a la OCDE, con una correlación positiva entre rotación laboral y crecimiento económico, sugiere que la IA puede ser un motor de desarrollo, sin embargo, estas tendencias muestran que esto solo será efectivo si se acompaña de políticas que mitiguen la exclusión y promuevan una transición inclusiva. Estos halagos sientan las bases para las conclusiones, donde se discutirá el equilibrio entre progreso económico y equidad social.

Por otra parte, se destaca la importancia del rol de la institucionalidad en la adopción de la IA desde una visión un poco más neutral. Las sociedades con instituciones sólidas que promuevan políticas equitativas y formación adecuada liderarán los beneficios de la IA, mientras que una gestión deficiente podría exacerbar desigualdades. Esta tendencia subraya la importancia de la gobernanza para equilibrar innovación y equidad.

## **Conclusiones**

En este estudio, se analizó en qué medida los impactos laborales de la IA configuran una condición para el desarrollo o, por el contrario, una forma de exclusión estructural en los mercados laborales contemporáneos. Los resultados identifican tres tendencias principales en relación con el riesgo y la sustitución laboral, tres tendencias vinculadas a la

creación de nuevos roles y al aumento de la productividad, y una tendencia neutral respecto al papel de la institucionalidad en la adopción de la IA. Estos resultados permiten comprender la dualidad de la IA como motor de transformación laboral y como desafío para la equidad en los mercados de trabajo, sentando así las bases para reflexionar sobre sus implicaciones sociales y económicas.

Dentro de los riesgos y la sustitución laboral, la primera tendencia identificada es la sustitución de empleos en múltiples niveles de cualificación. Si bien Wang et al. (2024) advierte que todavía es demasiado pronto para cuantificar los efectos de la IA en el mercado laboral, es innegable que la IA está sustituyendo empleos (Frey & Osborne, 2017), y no solo, como comúnmente se cree, en empleos de baja cualificación, sino también en roles cognitivos (Suleyman & Bhaskar, 2023) y niveles gerenciales medios (Vomberg et al., 2024). Por lo tanto, la IA no debe ser considerada únicamente como una herramienta, sino como un agente capaz de tomar decisiones e inventar ideas (Harari, 2024). En este sentido, Bankis y Formosa (2023) advierten que la pérdida de roles puede deshumanizar el trabajo.

Otro aspecto crítico de los riesgos y la sustitución laboral es el impacto desigual de la IA según la demografía, la geografía y los sectores. Huang (2024) se refiere a la heterogeneidad del impacto de la IA en el mercado laboral, sugiriendo la existencia de una jerarquización en cuanto a la

vulnerabilidad en el empleo, basada en factores sociodemográficos, lo que deja a ciertos grupos más expuestos a la pérdida de la fuente de dignidad e identidad que proporciona el trabajo (Arendt, 2013).

Complementando estos riesgos, una tercera tendencia en la sustitución laboral se centra en el desplazamiento de trabajadores, las nuevas demandas y las fricciones en el mercado laboral. Frey y Osborne (2017) destacan que el desplazamiento puede mitigarse a través de la educación y la capacitación en áreas prioritarias, como las correspondientes a STEM, lo que generaría nuevos nichos de mercado. Sin embargo, estas nuevas demandas tecnológicas pueden dar lugar a fricciones salariales, y las empresas podrían optar por automatizar más tareas en lugar de ajustar salarios.

Una de las principales tendencias en la creación de nuevos roles y el aumento de la productividad es la generación de nuevos empleos y habilidades impulsados por la IA. El fomento de habilidades en IA representa una ventaja competitiva para los jóvenes, permitiéndoles posicionarse en el mercado laboral (Cazzaniga, 2024). Además, la adopción de la IA en roles establecidos generará un efecto de reinstalación en empleos de ciberseguridad, salud y marketing digital (Hampole et al., 2025). En este sentido, Bankins y Formosa (2023) señala que la creación de nuevos roles podría mitigar

el riesgo de deshumanización del trabajo si la transformación se centra en tareas significativas.

A partir de la emergencia de nuevos roles, una segunda tendencia destaca la complementariedad entre humanos e IA como un factor clave para mejorar la productividad. En lugar de reemplazar, la IA complementa el trabajo a través de la relación triádica consumidor-tecnología-trabajador, que permite a los empleados priorizar la atención al consumidor mientras derivan tareas rutinarias a tecnologías autónomas (Vomberg et al., 2024). Esta colaboración no solo optimiza los procesos, sino que también potencia la capacidad de los trabajadores para generar valor en sus funciones.

Complementando esta tendencia, una tercera se centra en la reducción de desigualdades y la promoción de una transición inclusiva. Autor et al., (2023) enfatiza que la adopción de la IA es un proceso lento, lo que ofrece tiempo para preparar a los trabajadores y ajustar las instituciones. En este contexto, la IA representa una oportunidad para redefinir el trabajo, si se priorizan las capacidades humanas (Danaher, 2019), fomentando así un mercado laboral más equitativo y productivo.

Entre las tendencias de riesgos y sustitución laboral, y las de creación de nuevos roles y aumento de productividad, coexiste una posición neutral que no se enfoca en aspectos favorables o desfavorables, sino en la administración de las

oportunidades de innovación que proporciona la IA a través de las instituciones políticas y económicas.

Para que la IA beneficie al mercado laboral global, los tomadores de decisiones deben fortalecer instituciones inclusivas que promuevan la educación, la regulación equitativa y la distribución de beneficios. En ausencia de estas, la IA podría amplificar desigualdades, replicando los patrones de fracasos descritos por Acemoglu y Robinson (2013). Por lo tanto, al reflexionar sobre los responsables de los impactos negativos o positivos de la IA en los valores sociales (Santoni de Sio, 2024) será necesario replantear el concepto de institucionalidad propuesto por Acemoglu y Robinson (2013).

El análisis de las tendencias revela que el impacto de la IA en el mercado laboral no debe centrarse únicamente en los efectos observables, sino también en el comportamiento previo de las instituciones, el cual desempeña un papel crucial en la configuración de resultados positivos o negativos. Este estudio demuestra que la IA, lejos de ser únicamente una fuente de desplazamiento laboral, genera más empleos de los que elimina, especialmente en roles técnicos, pero también en cargos que requieren habilidades cognitivas avanzadas. Así, la IA emerge como una oportunidad para el crecimiento económico, siempre que las instituciones promuevan un entorno inclusivo que mitigue las dinámicas de exclusión y potencie la creación de empleo.

A pesar de los aportes de este estudio, es importante reconocer algunas limitaciones. En primer lugar, la incertidumbre en la cuantificación de los impactos de la IA, como señala Wang et al., (2024) restringe la capacidad de prever con precisión la magnitud de la sustitución y creación de empleos, especialmente en contextos dinámicos y emergentes. En segundo lugar, el análisis se centra en mercados laborales de economías desarrolladas, lo que limita su aplicabilidad a economías en desarrollo, donde las dinámicas institucionales y tecnológicas pueden diferir significativamente. Finalmente, la falta de datos longitudinales impide evaluar los efectos a largo plazo de la adopción de la IA, lo que representa una oportunidad para futuras investigaciones que aborden estas brechas.

## Referencias

- Acemoglu, D., Autor, D., & Johnson, S. (2023). Can we have pro-worker AI? Choosing a path of machines in service of minds. *CEPR Policy Insight*, 123.
- Acemoglu, D., & Johnson, S. (2023). *Power and progress: Our thousand-year struggle over technology and prosperity*. PublicAffairs.
- Acemoglu, D., & Johnson, S. (2024). Learning from Ricardo and Thompson: Machinery and labor in the early industrial revolution and in the age of artificial intelligence. *Annual Review of Economics*, 16, 597–

621. <https://doi.org/10.1146/annurev-economics-091823-025129>

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020). Competing with robots: Firm-level evidence from France. *AEA Papers and Proceedings*, 110, 383–388.

Acemoglu, D., & Restrepo, P. (2018). *Artificial intelligence, automation and work* (NBER Working Paper No. 24196). National Bureau of Economic Research. <https://doi.org/10.3386/w24196>

Acemoglu, D., & Robinson, J. A. (2013). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Currency.

Arendt, H. (2013). *The human condition* (2nd ed.). University of Chicago Press.

Bankins, S., & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics*, 185, 725–740. <https://doi.org/10.1007/s10551-023-05339-7>

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of Economic Research. <https://doi.org/10.3386/w31161>

Cazzaniga, M. (2024). Gen-AI: Artificial Intelligence and the Future of Work. *IMF Staff Discussion Notes*, 001.

Coles, M. G., & Mortensen, D. T. (2016). Equilibrium labor turnover, firm growth, and unemployment. *Econometrica*, 84(1), 347–363.



- Danaher, J. (2019). *Automation and utopia: Human flourishing in a world without work*. Harvard University Press.
- Deming, D., Ong, C., & Summers, L. H. (2024). Technological disruption in the US labor market. En M. S. Kearney & L. Pardue (Eds.), *Strengthening America's economic dynamism*. Aspen Institute.
- Demirci, O., Hannane, J., & Zhu, X. (2025). Who is AI replacing? The impact of generative AI on online freelancing platforms. *Management Science*.
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Hampole, M., Papanikolaou, D., Schmidt, L. D. W., & Seegmiller, B. (2025). *Artificial intelligence and the*

- labor market* (NBER Working Paper No. 33509). National Bureau of Economic Research. <https://doi.org/10.3386/w33509>
- Harari, Y. N. (2017). Reboot for the AI revolution. *Nature*, 550(7676), 324–327. <https://doi.org/10.1038/550324a>
- Harari, Y. N. (2024). *Nexus: A brief history of information networks from the stone age to AI*. Harper.
- Huang, Y. (2024). *The labor market impact of artificial intelligence: Evidence from US regions* (IMF Working Paper No. 2024/199). International Monetary Fund.
- Janzen, M., & DeRoo, N. (2022). What is the value of a person when artificial intelligence can do all the work? En C. El Morr (Ed.), *AI and society*. Chapman and Hall/CRC.
- Machovec, C., Rieley, M. J., & Rolen, E. (2025). Incorporating AI impacts in BLS employment projections: Occupational case studies. *Monthly Labor Review*. <https://doi.org/10.21916/mlr.2025.1>
- Santoni de Sio, F. (2024). Artificial intelligence and the future of work: Mapping the ethical issues. *Journal of Ethics*, 28, 407–427. <https://doi.org/10.1007/s10892-024-09493-6>
- Sen, A. (1999). *Development as freedom*. Oxford University Press.
- Shen, Y., & Zhang, X. (2024). The impact of artificial intelligence on employment: The role of virtual agglomeration. *Humanities and Social Sciences*

*Communications*, 11, 122.

<https://doi.org/10.1057/s41599-024-02647-9>

Siebert, W. S., & Zubanov, N. (2009). Searching for the optimal level of employee turnover. *Academy of Management Journal*, 52(2), 294–313.

<https://doi.org/10.5465/amj.2009.37308149>

Song, F., Agarwal, A., & Wen, W. (2023). The impact of generative AI on collaborative open-source software development: Evidence from GitHub Copilot. *SSRN*.

<https://doi.org/10.2139/ssrn.4856935>

Suleyman, M., & Bhaskar, M. (2023). *The coming wave: Technology, power, and the 21st century's greatest dilemma*. Random House.

Vomberg, A., de Haan, E., Fabian, N. E., & Broekhuizen, T. (2024). Digital knowledge engineering for strategy development. *Journal of Business Research*, 177, 114632.

<https://doi.org/10.1016/j.jbusres.2024.114632>

Wang, X., Chen, M., & Chen, N. (2024). How artificial intelligence affects the labour force employment structure. *Heliyon*, 10(5), e26686.

<https://doi.org/10.1016/j.heliyon.2024.e26686>

# **Razón pública y fragmentación epistémica: una defensa liberal frente a la crisis de la verdad pública**

*Public Reason and Epistemic Fragmentation: A Liberal Defense against the Crisis of Public Truth*

**Nicolás Fuentes Valdebenito<sup>8</sup>**

## **Resumen**

Este capítulo sostiene que las democracias liberales atraviesan una crisis de legitimidad epistémica: la erosión de las condiciones públicas de confianza, deliberación y justificación recíproca compromete los fundamentos sobre los que descansa la justificabilidad del poder ante ciudadanos

---

<sup>8</sup> Centro Interdisciplinario en Ética, Política y Economía, Chile.  
<https://orcid.org/0000-0002-2229-7488>

libres e iguales. Frente a la posverdad, la desinformación algorítmica y la tribalización digital, examina la tentación de responder mediante una custodia estatal o tecnocrática de la verdad y argumenta que tales remedios son autodestructivos, pues pretenden producir coercitivamente una confianza que la coerción no puede producir. Propone, como alternativa, un *liberalismo epistémico*: una reconstrucción de las condiciones estructurales de la confianza pública que protege la arquitectura del espacio público sin imponer sus contenidos, y lo proyecta sobre el problema de la gobernanza algorítmica de la esfera pública.

**Palabras clave:** Legitimidad epistémica; razón pública; confianza pública; desinformación; gobernanza algorítmica.

## **Abstract**

This chapter argues that liberal democracies face a crisis of epistemic legitimacy: the erosion of the public conditions of trust, deliberation, and reciprocal justification compromises the foundations on which the justifiability of power to free and equal citizens rests. Against post-truth, algorithmic disinformation, and digital tribalization, it examines the temptation to respond through state or technocratic guardianship of truth and argues that such remedies are self-defeating, since they attempt to produce coercively a trust that coercion cannot produce. As an alternative, it proposes epistemic liberalism: a reconstruction

of the structural conditions of public trust that protects the architecture of the public sphere without imposing its contents, and brings this view to bear on the problem of the algorithmic governance of the public sphere.

**Keywords:** Epistemic legitimacy; public reason; public trust; disinformation; algorithmic governance.

### **Introducción: una variedad olvidada de crisis de legitimidad**

Las democracias liberales se fundan sobre una promesa exigente: que el ejercicio del poder coercitivo pueda justificarse ante quienes lo padecen mediante razones que, en principio, podrían aceptar. Esta idea, que recorre la tradición que va de Locke a Rawls, supone que la legitimidad no descansa únicamente en procedimientos formales ni en resultados eficientes, sino en una relación de justificación recíproca entre iguales. Como ha insistido Forst (2012), el núcleo de esa relación es un *derecho a la justificación*: la pretensión de no ser sometido a normas que no puedan justificarse con razones que uno no pueda razonablemente rechazar. Lo que rara vez se explicita es que tal relación presupone condiciones epistémicas. Para que las razones públicas circulen, los ciudadanos deben compartir prácticas que les permitan distinguir afirmaciones fundadas de meras aserciones, reconocer fuentes dignas de crédito y revisar sus creencias a la luz de la evidencia. La justificación política se

sostiene sobre una infraestructura epistémica común, tan poco visible como el aire y, como el aire, advertida sobre todo cuando empieza a faltar.

La tesis que aquí se desarrolla es que esa infraestructura se encuentra hoy erosionada, y que tal erosión constituye una crisis de legitimidad de un tipo particular, que propongo denominar *crisis de legitimidad epistémica*. La precisión importa, porque la literatura sobre la “crisis de la democracia” amalgama fenómenos heterogéneos. Conviene distinguir tres variedades. Una *crisis de eficacia* sobreviene cuando las instituciones dejan de producir los bienes que se esperan de ellas; una *crisis de representación*, cuando los ciudadanos dejan de reconocerse en quienes deciden por ellos. La que aquí me ocupa es distinta y más básica: una crisis de las condiciones bajo las cuales una controversia pública puede todavía tratarse como susceptible de resolución mediante razones. No se trata de que circulen más falsedades que antes, la mentira política es tan antigua como la política, y ya Arendt (2006) advertía sobre la fragilidad perenne de la verdad factual frente al poder, sino de que se debilitan las condiciones que permiten que la corrección de una falsedad sea reconocida como tal por quienes la sostienen. Cuando la posverdad, la desinformación a escala industrial, la tribalización algorítmica y la desconfianza institucional se combinan, lo comprometido no es esta o aquella creencia, sino la confianza de fondo que hace posible la deliberación.

Frente a este diagnóstico, una respuesta cada vez más audible propone que el Estado, o una casta de expertos y de plataformas reguladas, asuma la tarea de custodiar la verdad pública: certificar lo verdadero, suprimir lo falso, reorientar el espacio informativo hacia resultados epistémicamente correctos. Llamaré a esta familia de propuestas, con deliberada amplitud, *perfeccionismo epistémico*: la tesis de que es legítimo que el poder político promueva creencias verdaderas y suprima creencias falsas en nombre del bien de los ciudadanos o de la calidad de la democracia.

Su atractivo es comprensible: si el problema es epistémico, parece sensato confiar su solución a quienes saben.<sup>9</sup> La hipótesis central es que esa vía resulta autodestructiva, pues pretende restaurar la legitimidad sacrificando aquello que la legitimidad liberal protege: la libertad de conciencia, el pluralismo de las concepciones del bien y el derecho de cada ciudadano a ser tratado como fuente de razones, no como objeto de gestión. Quien cura la enfermedad matando al paciente no ha curado nada.

---

<sup>9</sup> Empleo “perfeccionismo epistémico” en sentido amplio: abarca el perfeccionismo moral clásico (la promoción estatal de concepciones del bien) y su análogo cognitivo, la promoción coactiva de creencias verdaderas y la supresión de creencias falsas en nombre del bien de los ciudadanos. La continuidad estructural entre ambos es lo que el argumento pretende exhibir.



La tesis puede enunciarse, pues, con precisión: la crisis contemporánea de la verdad pública es una crisis de legitimidad epistémica porque erosiona las condiciones de justificación recíproca necesarias para el liberalismo político; pero la respuesta no puede consistir en un perfeccionismo epistémico, estatal o tecnocrático, sino en una reconstrucción liberal de las condiciones estructurales, institucionales y éticas de la confianza pública. Tal reconstrucción sostiene dos compromisos que suelen presentarse como incompatibles. Por un lado, tomar en serio que la legitimidad democrática depende parcialmente de condiciones epistémicas compartidas: no toda configuración del espacio informativo es indiferente para la justificación. Por otro, resistir la deriva perfeccionista: reconocer la importancia política de esas condiciones no autoriza al Estado a imponer una concepción sustantiva de la verdad. El liberalismo que aquí se defiende es, por ello, *epistémicamente sensible pero no perfeccionista*, y gira en torno a una distinción decisiva: la que media entre proteger las condiciones del juicio y sustituir el juicio.

La contribución del capítulo conviene declararla sin rodeos, pues la teoría política contemporánea ha producido ya diagnósticos penetrantes de la crisis epistémica, de Rauch (2021) a Lynch (2016), de Sunstein (2017) a Samaržija y Cassam (2023), y un creciente cuerpo de trabajo sobre la gobernanza algorítmica de lo público (Lazar, 2025; Douek, 2021). Lo que aquí se ofrece es una articulación sistemática que (i) identifica la dimensión epistémica de la legitimidad

como una exigencia normativa diferenciada; (ii) muestra que las respuestas perfeccionistas (paternalista, epistocrática y algorítmico-tutelar) comparten una misma estructura inválida; y (iii) deriva un criterio operativo (un *test de doble condición*) capaz de distinguir las intervenciones liberalmente admisibles sobre la esfera pública de las que no lo son. El concepto de *liberalismo epistémico* nombra el lugar teórico que resulta de sostener esas tres tesis: ni indiferencia ante la arquitectura del espacio público, ni tutela de sus contenidos. Inscrito en los debates sobre epistemología social, ética de la información y tecnopolítica que organizan este volumen, el capítulo busca ofrecer un criterio normativo común a problemas hoy dispersos entre la teoría de la democracia y la ética de las plataformas.

## **La crisis epistémica como crisis de legitimidad**

El problema epistémico de las democracias contemporáneas no se reduce a que circule más información falsa; si así fuera, bastaría con multiplicar los verificadores y educar mejor a la ciudadanía. El fenómeno es más profundo: se han debilitado las condiciones que permiten que la corrección de una falsedad sea reconocida como tal por quienes la sostienen. La desinformación madura no opera tanto persuadiendo de proposiciones falsas cuanto erosionando la confianza en los mecanismos sociales de validación del conocimiento, como la ciencia, el periodismo profesional, las instituciones de control, de modo que deje de

existir una instancia compartida ante la cual una afirmación pueda ser refutada. Rauch (2021) ha caracterizado ese orden como una *constitución del conocimiento*: una red de reglas e instituciones (revisión por pares, contrastación empírica, libertad de crítica, separación entre la verdad de una afirmación y la identidad de quien la formula) que convierte el desacuerdo en motor de descubrimiento y no en guerra de facciones. Lo distintivo de la crisis actual es que esa constitución no es atacada negando hechos particulares, sino socavando la idea misma de que existan procedimientos impersonales para establecerlos. La estrategia ya no es la mentira, sino el cinismo: si toda fuente responde a un interés, no hay diferencia relevante entre el experto y el charlatán, y la elección entre versiones del mundo se vuelve cuestión de lealtad tribal antes que de evidencia.

La distinción más fértil para comprender el proceso la debemos a Nguyen (2020), entre la *burbuja epistémica* y la *cámara de eco*. La burbuja omite las voces contrarias, por los algoritmos de recomendación o por la homofilia de las redes, dejamos de exponernos a quienes piensan distinto, y es frágil, pues basta exponer al sujeto a la evidencia omitida para perforarla. La cámara de eco, en cambio, *desacredita activamente* las fuentes externas: inculca razones para desconfiar sistemáticamente del disidente y reinterpreta toda evidencia contraria como confirmación de la conspiración. De ahí que resista a la mera información: el problema no es un déficit de datos, sino una manipulación de la confianza. Esta

consecuencia gobernará el argumento ulterior: las soluciones puramente informacionales son impotentes frente a patologías que son, en su raíz, patologías de la confianza. La evidencia empírica reciente recomienda cautela ante los diagnósticos catastrofistas: la prevalencia efectiva de la desinformación suele ser menor que la alarma que suscita; pero ello no debilita el punto normativo: aun una incidencia modesta puede tener efectos desproporcionados cuando opera erosionando la confianza en las instancias de validación.

Esta fragmentación no es un accidente cultural ni mero resultado de vicios individuales: tiene una base infraestructural que la teoría política no puede ignorar sin volverse ingenua. Las plataformas que median hoy buena parte de la comunicación pública no están diseñadas para optimizar la calidad epistémica del debate, sino para maximizar el *engagement*. Y como los contenidos que suscitan reacciones emocionales intensas retienen mejor la atención, los sistemas de recomendación amplifican aquello que degrada la deliberación: no por malicia, sino por la lógica de un modelo de negocio en que la atención es el recurso escaso. Sunstein (2017) ha mostrado cómo esta arquitectura favorece la *polarización de grupo* y cómo la fragmentación de la esfera pública en nichos sin solapamiento erosiona la experiencia compartida y la serendipia: esa exposición no buscada a perspectivas que uno no habría seleccionado, condición de una ciudadanía capaz de razonar junta. Lynch

(2016) ha capturado la paradoja con una fórmula precisa: en la era de los datos masivos sabemos más y comprendemos menos.

Han ha radicalizado el diagnóstico al señalar una mutación del poder. Si en la sociedad disciplinaria el poder operaba por la prohibición y la vigilancia externa, en lo que Han (2014) llama la *psicopolítica* digital opera de manera más sutil: no reprime, seduce; no niega la libertad, la explota. El sujeto contemporáneo no es coaccionado a exhibirse y consumir: lo hace voluntariamente, experimentando como libertad lo que es sujeción a la valorización de los datos. En su diagnóstico más reciente, Han (2022) describe una *infocracia* en que el régimen de la información sucede al disciplinario: el poder ya no censura, satura; ya no silencia, distrae. La amenaza a la autonomía epistémica no proviene principalmente de la censura estatal, sino de una arquitectura que modela las preferencias desde dentro, presentándose como mera satisfacción de deseos preexistentes. El enemigo de la autonomía ya no es solo el Leviatán que prohíbe, sino también el dispositivo que adula, un entorno que, como ha observado Cassam (2019), recompensa los *vicios epistémicos*: la credulidad militante, el cierre cognitivo, el pensamiento desiderativo. Conviene, con todo, resistir la tentación de leer a Han como si clausurara la agencia: que la arquitectura module las preferencias no implica que las determine.

Resta mostrar por qué este diagnóstico, que podría parecer asunto de epistemólogos, es una cuestión de *legitimidad* política y no solo de higiene cognitiva. La conexión es esta: si la legitimidad liberal consiste en la justificabilidad del poder ante ciudadanos que pueden reconocer las razones que se les ofrecen, presupone entre ellos un acervo mínimo de hechos no disputados, de fuentes mutuamente reconocibles y de normas compartidas de argumentación. Cuando ese suelo se fractura, no se discute ya desde marcos distintos sobre un mundo común, sino desde mundos distintos que no se tocan, y la justificación recíproca que Forst (2012) sitúa en el centro de la legitimidad deja de tener destinatario posible. Williams (2002) lo vio desde otra dirección: cuando una cultura “desenmascara” las virtudes de la verdad como simples máscaras del poder, no se libera de ellas, sino que se priva del lenguaje en que podría denunciar el engaño. La crisis epistémica es, en este sentido preciso, una crisis política: no porque haya falsedades, siempre las hubo, sino porque amenaza las condiciones bajo las cuales la verdad y la falsedad pueden todavía importar colectivamente.

Reconocer la dimensión infraestructural del problema tensiona, sin embargo, la propia posición liberal. Si las condiciones epistémicas de la ciudadanía dependen de arquitecturas diseñadas por actores privados con fines comerciales, la indiferencia del Estado liberal hacia ellas no es neutralidad, sino ratificación tácita de un orden epistémico configurado por el mercado de la atención: el liberalismo no

puede refugiarse en un *laissez-faire* informacional como si el statu quo fuera inocente. Pero tampoco puede saltar de esa constatación a la conclusión de que el Estado deba apoderarse del espacio informativo. Entre la indiferencia y el control se extiende un espacio normativo que la teoría liberal está obligada a cartografiar, y cuyo trazado constituye la tarea positiva de este capítulo.

### **¿Por qué fracasa la custodia de la verdad?**

Para evaluar las respuestas posibles es preciso reconstruir qué entiende el liberalismo por legitimidad. El núcleo puede formularse mediante lo que Rawls llamó el *principio liberal de legitimidad*. En su formulación canónica,<sup>10</sup> “nuestro ejercicio del poder político es plenamente correcto solo cuando se ejerce de acuerdo con una constitución cuyos elementos esenciales cabe esperar razonablemente que todos los ciudadanos, en cuanto libres e iguales, suscriban” (Rawls, 2005a, p. 137). La legitimidad es, así, una relación de justificabilidad: no basta con que las decisiones se tomen por procedimientos democráticos; deben poder justificarse ante cada ciudadano mediante razones que este pueda, en principio, reconocer como tales. El acento no recae sobre la corrección del resultado ni sobre la regularidad

---

<sup>10</sup> Salvo indicación contraria, las traducciones de las obras citadas en lengua extranjera son responsabilidad del autor.

del procedimiento, sino sobre la posibilidad de rendir cuentas a cada uno como a un igual.

Esta exigencia se vuelve apremiante por el *hecho del pluralismo razonable* (Rawls, 2005a): en condiciones de libertad, los ciudadanos sostienen de manera persistente doctrinas comprensivas distintas e incompatibles, sin que ello implique irracionalidad. Rawls lo atribuyó a las *cargas del juicio*: la complejidad de la evidencia, los pesos distintos asignados a consideraciones igualmente pertinentes, la indeterminación de los conceptos, la diversidad de las experiencias que configuran nuestras valoraciones. Las cargas del juicio explican por qué el desacuerdo puede ser razonable: no toda discrepancia se debe a error, ignorancia o mala fe.

Aquí surge la primera tensión que el problema epistémico introduce en la teoría liberal. La doctrina de las cargas del juicio fue concebida para el desacuerdo sobre el valor; pero la crisis de la verdad pública involucra crecientemente el desacuerdo sobre *hechos*: si una vacuna es segura, si una elección fue fraudulenta, si el clima se calienta por causas antrópicas. ¿Es el desacuerdo factual también un desacuerdo razonable que el liberalismo deba tolerar en pie de igualdad con el valorativo? Sostendré que no.<sup>11</sup> Hay una

---

<sup>11</sup> Reservo “desacuerdo razonable” para la divergencia valorativa que sobrevive al ejercicio escrupuloso de la razón, y “desacuerdo factual” para la divergencia sobre cuestiones de hecho susceptibles, en principio, de



diferencia normativa fundamental entre el pluralismo sobre el bien y el “pluralismo” sobre los hechos; ignorarla conduce a un relativismo que el liberalismo no necesita ni debe asumir. La distinción no es tajante: la literatura sobre el *desacuerdo profundo* ha mostrado que las controversias factuales suelen estar cargadas de compromisos valorativos sobre qué cuenta como evidencia (Ranalli & Lagewaard, 2022); pero su porosidad no autoriza a colapsarla. Como argumenta Lynch (2012), una democracia no puede prescindir de patrones de evaluación comunes sin disolverse en una contienda de poder: defender la objetividad de ciertos hechos no es autoritarismo epistémico, sino la condición de que el propio desacuerdo valorativo sea siquiera inteligible.

La asimetría conduce al problema de la autoridad epistémica. Si no todas las creencias factuales son equivalentes, hay quienes saben más sobre ciertas materias. Goldman (2001) planteó la dificultad del *lego*: ¿cómo evaluar cuál de dos expertos en desacuerdo es más digno de crédito? Señaló diversas fuentes de evidencia indirecta: la calidad dialéctica de los argumentos, el acuerdo de terceros, las credenciales y el historial de aciertos, la presencia de intereses; y el punto decisivo es que la deferencia racional al experto no es una abdicación de la autonomía, sino su ejercicio. Conviene recordar, con Rosenfeld (2018), que la

---

resolución mediante procedimientos públicos de validación. La frontera es porosa, pero su porosidad no la disuelve.

tensión entre el saber experto y el juicio del ciudadano no es una patología de la era digital, sino una ambigüedad inscrita en la democracia desde sus orígenes. Pero la deferencia racional presupone lo que la crisis ha erosionado: la confianza en que existen expertos identificables y en que las credenciales señalan competencia. Cuando la cámara de eco ha inculcado la sospecha de que toda credencial es captura, los marcos goldmanianos se vuelven inoperantes. Esta es, en clave epistemológica, la lógica del *populismo epistémico*: la valorización del saber del “pueblo llano” frente a un establishment cognitivo presentado como duplicidad organizada (Saurette & Gunster, 2011), posición que no combate tal o cual afirmación experta, sino la división misma del trabajo epistémico.

Fricker (2007) introduce una dimensión que el argumento no puede ignorar. Mostró que existe una *injusticia epistémica* testimonial, “un mal que se le hace a alguien específicamente en su capacidad de sujeto de conocimiento” (p. 1), que ocurre cuando los prejuicios del oyente otorgan a un hablante menos credibilidad de la que merece. Su relevancia es ambivalente. Por un lado, advierte que la apelación a la “autoridad epistémica” puede encubrir exclusiones, y Medina (2013) ha extendido el análisis hasta una *epistemología de la resistencia* que muestra cómo los privilegiados desarrollan vicios (arrogancia, ignorancia activa) que distorsionan las economías de credibilidad. Por otro, el concepto se ha invocado para deslegitimar toda

jerarquía de credibilidad como mera dominación, lo que llevaría a un nivelamiento igualmente corrosivo; y es revelador que el propio Medina persiga una redistribución más justa de la credibilidad, no su abolición. Una concepción liberal defendible sostiene ambas cosas: que las jerarquías de credibilidad pueden ser injustas y deben criticarse, y que no por ello toda jerarquía lo es. La existencia de la falsificación de moneda no prueba que el dinero sea una ilusión.

Si la crisis amenaza la legitimidad, ¿por qué no habría de corresponder al Estado proteger el entorno epistémico? La respuesta liberal exige distinguir dos versiones clásicas de la propuesta: la paternalista y la tecnocrática; a las que el desarrollo tecnológico ha añadido una tercera, algorítmica, que trataré luego. Las tres comparten la estructura del perfeccionismo epistémico. El paternalismo informacional sostiene que, siendo los ciudadanos cognitivamente falibles, el Estado puede protegerlos de sus propios errores, sea suprimiendo contenidos falsos, sea mediante técnicas de orientación cognitiva. Su versión sofisticada es el *paternalismo libertario* o la política de los *nudges* de Thaler y Sunstein (2008): puesto que toda arquitectura de elección influye sobre las decisiones, el diseñador podría orientar al ciudadano hacia mejores opciones sin coaccionarlo; aplicado al dominio epistémico, esto sugeriría diseñar el entorno informativo para empujar hacia las fuentes confiables.

La objeción liberal no es que la falibilidad cognitiva sea irreal (lo es) ni que toda arquitectura de elección pueda evitarse (no puede), sino algo más fundamental, que concierne a la relación de respeto entre el Estado y el ciudadano. Tratar a los ciudadanos como sujetos cuyas creencias deben ser gestionadas por una autoridad que sabe mejor lo que les conviene creer es incompatible con tratarlos como agentes racionales autónomos. Hay una diferencia moral categórica entre ofrecer razones a un agente y manipular las condiciones bajo las cuales forma sus creencias: lo primero respeta su autonomía; lo segundo la elude. Y, como argumentó O'Neill (2002), los regímenes de control destinados a *forzar* la confiabilidad tienden a corroer la confianza que dicen perseguir, porque comunican sospecha donde pretenden generar seguridad. A esta objeción moral se suma una epistémica: el paternalista debe presuponer un acceso privilegiado a la verdad en disputa, y esa presuposición es la que la falibilidad invocada para justificar la intervención vuelve insostenible. Si los ciudadanos pueden equivocarse, también pueden hacerlo los funcionarios y los algoritmos que decidirían qué es desinformación. El falibilismo, tomado en serio, milita contra el perfeccionismo: precisamente porque podemos equivocarnos, ninguna instancia debe arrogarse el poder de clausurar la controversia.

La segunda versión es tecnocrática. Ante la presunta incompetencia epistémica del *demos*, propone trasladar el poder de decisión a quienes poseen competencia. Su

formulación más franca es la *epistocracia* de Brennan (2016): si el conocimiento político está desigualmente distribuido y las malas decisiones colectivas dañan a todos, podría ser legítimo ponderar el poder en favor de los más competentes. El argumento tiene fuerza y no debe despacharse con indignación, sino con razones, que proceden en tres pasos. El primero, la *demografía del conocimiento*: cualquier criterio para identificar a los competentes seleccionará a ciertos grupos y excluirá a otros, de modo que la epistocracia institucionalizaría los sesgos de los dominantes bajo apariencia de competencia neutral, la injusticia epistémica elevada a principio constitucional. El segundo, la *competencia ilusoria*: el saber experto es local y de dominio específico; no existe una experticia general en “lo político” que reúna la epidemiología, la macroeconomía y la teoría de la justicia.

El tercer paso, el más fundamental, es la objeción de la *legitimidad*, y aquí Estlund (2008) ofrece el argumento decisivo. Aun si pudiéramos identificar de modo incontrovertible a los más competentes, de su mayor saber no se seguiría su derecho a gobernar. Es la falacia del *experto devenido jefe*: que alguien sepa mejor que yo lo que me conviene no lo convierte en mi jefe; del valor de verdad de una proposición no se sigue autoridad política sobre quienes la rechazan. Y la razón es estrictamente liberal: la identificación de los expertos está ella misma sujeta a desacuerdo razonable entre puntos de vista cualificados. Forst (2012) ilumina la

misma conclusión: tanto el censor que suprime como el epistócrata que gobierna en lugar de los demás niegan a las personas su condición de iguales en el espacio de las razones. En una línea convergente, Urbinati (2014) ha mostrado que reducir la democracia a un procedimiento de producción de decisiones correctas (su “desfiguración” epistémica) traiciona su sentido como autogobierno de iguales. Es notable que epistocracia y populismo, presentados como antagonistas, compartan la estructura aquí rechazada: ambos sustituyen la pregunta por el derecho a decidir por una pretensión de acceso privilegiado a la verdad, la competencia del experto en un caso, la autenticidad del pueblo en el otro.

Cabe formular una conclusión parcial. Paternalismo y tecnocracia fracasan por la misma razón profunda: ambos pretenden resolver un déficit de confianza mediante un ejercicio de poder, sin advertir que la confianza, por su naturaleza, no puede producirse coercitivamente. Lo que el poder produce es sumisión o resistencia, no confianza. Y como la legitimidad epistémica de la democracia depende de condiciones de confianza, los remedios perfeccionistas erosionan justamente lo que pretenden restaurar.

### **La reconstrucción liberal: razón pública, confianza y virtudes**

Si las soluciones perfeccionistas están descartadas, ¿qué recursos ofrece la propia tradición liberal? Sostengo que

se articulan en torno a tres nociones que se reclaman mutuamente: la razón pública como ética del discurso cívico, la confianza pública como bien relacional irreductible al control, y las virtudes epistémicas cívicas como disposiciones de los ciudadanos.

La idea rawlsiana de razón pública suele malinterpretarse como una restricción de lo que los ciudadanos pueden decir. Pero, bien entendida, no es una mordaza jurídica, sino un *deber moral de civilidad*: el compromiso de justificar las propias posiciones sobre las cuestiones constitucionales fundamentales con razones que otros, partiendo de concepciones del bien diferentes, puedan razonablemente aceptar (Rawls, 2005b). Es una ética, no una ley. Rawls precisó, además, que no obliga a excluir las doctrinas comprehensivas del debate, sino que admite introducirlas siempre que, a su debido tiempo, se ofrezcan también razones propiamente públicas que las respalden: la *cláusula restrictiva*. Habermas (1996) ofreció una articulación complementaria desde la ética del discurso: la legitimidad de las normas depende de que puedan ser aceptadas por todos los afectados en un discurso libre de coacción, donde no opere más fuerza que la “coacción no coactiva del mejor argumento”. En su intercambio con Rawls, Habermas (1995) objetó que la prioridad de “lo razonable” subordina la verdad a la mera aceptabilidad; Rawls (1995) replicó que el liberalismo político es una concepción *independiente*, y que la prioridad de lo razonable es la

condición de una cooperación equitativa entre quienes disienten sobre la verdad última. Por debajo de la disputa, ambos conciben la justificación como un proceso regido por normas internas: reciprocidad, sinceridad, disposición a dar y pedir razones; que Gutmann y Thompson (1996) hicieron principio cardinal de la democracia deliberativa. La crisis epistémica es, desde este ángulo, una crisis de tales normas; y ese compromiso no se restaura por decreto.

Conviene incorporar una crítica interna que impide que la razón pública degenera en coartada de la deferencia. Lafont (2020) ha objetado a las lecturas que inducen a los ciudadanos a aceptar como propias políticas que no pueden examinar, delegando ciegamente en expertos por la vía de un “atajo” que les ahorra el trabajo de comprender. Su democracia deliberativa exige que los ciudadanos puedan en todo momento demandar razones. La objeción es justa: deferir al climatólogo sobre la magnitud del calentamiento es racional, pero qué hacer al respecto (cómo distribuir los costos de la mitigación) es una cuestión irreductiblemente sujeta a la deliberación de todos. Esta advertencia resultará decisiva cuando el “atajo” deje de ser el del experto y pase a ser el del algoritmo.

El concepto que articula la propuesta positiva es el de *confianza pública*, que conviene distinguir de la mera credulidad. Siguiendo a O'Neill (2002), la confianza bien depositada no es un sentimiento difuso, sino un juicio



diferenciado sobre la *confiabilidad* de agentes e instituciones en dominios específicos: confío en mi médico para cuestiones médicas, no para asesoría financiera. De ahí una tesis normativa: el objetivo de una política epistémica liberal no es maximizar la confianza (una confianza indiscriminada sería desastrosa), sino fomentar las condiciones bajo las cuales los ciudadanos puedan otorgarla y retirarla con buen juicio. El foco se desplaza de la confianza demandada a la confiabilidad debida: antes que exigir a los ciudadanos que confíen, hay que exigir a las instituciones que la merezcan y se hagan inteligibles a un juicio capaz de discriminar.

Vallier (2019, 2021) ha desarrollado una teoría de la confianza especialmente pertinente para una sociedad plural. Distingue la confianza social, es decir, esperar comportamiento cooperativo de conciudadanos desconocidos; de la confianza política en las instituciones, y muestra que ambas dependen de que estas operen según reglas públicamente justificables. Su tesis central es que la confianza es compatible con el desacuerdo profundo, y de hecho lo presupone, siempre que exista una justificación pública de las reglas: quienes discrepan radicalmente sobre el bien pueden confiar unos en otros si comparten la convicción de que las instituciones los tratan según principios que cada cual tiene razón para aceptar. Vallier articula esa justificación en clave *convergente* antes que consensual: admite que cada ciudadano respalde las mismas reglas desde su propio conjunto de razones. La consecuencia es liberadora: la

confianza pública no exige homogeneidad de creencias, sino justificabilidad de las reglas. De aquí la tesis que vertebra la propuesta: la confianza pública es un bien relacional que ninguna autoridad puede producir directamente, pero que las instituciones pueden volver posible o imposible según cómo se comporten. El Estado no puede ordenar que se confíe en él, pero puede hacerse confiable. El perfeccionista trata la confianza como un *output* que fabricar; la política liberal de la confianza, como una respuesta libre a las condiciones que la merecen. No es, por tanto, un estado psicológico agregado que medir en encuestas, sino una categoría normativa y relacional.

El tercer recurso es ético en sentido estricto. Cortina (1997, 2017) ha desarrollado, en el marco de una ética cívica de mínimos, la idea de que las sociedades moralmente plurales pueden converger en ciertas virtudes públicas sin compartir una concepción comprehensiva del bien. Entre esas virtudes figuran disposiciones epistémicas decisivas (la veracidad, la disposición a escuchar a quien discrepa, la apertura a revisar las propias creencias, la responsabilidad sobre lo que se afirma y difunde) que Williams (2002) reconstruyó como las dos virtudes fundamentales de la verdad: la *exactitud*, cuidado por aquello que se cree, y la *sinceridad*, cuidado por aquello que se dice. Cortina advirtió, además, sobre la *aporofobia*, el rechazo al pobre, como un sesgo que distorsiona nuestras prácticas de atribución de credibilidad. La apelación a las virtudes podría parecer

ingenua: ¿no es justamente su erosión lo que constituye la crisis? Pero la objeción confunde el diagnóstico con la renuncia. Su cultivo no es empresa puramente exhortativa: tiene condiciones institucionales (la educación en el pensamiento crítico, los espacios deliberativos plurales, el periodismo profesional, la ciencia abierta). La diferencia, una vez más, separa cultivar las disposiciones que permiten juzgar bien de dictar qué debe juzgarse: lo primero capacita; lo segundo sustituye.

Estos tres recursos permiten formular la tesis intermedia que el capítulo defiende, equidistante del *laissez-faire* y del perfeccionismo: la legitimidad democrática depende de ciertas *condiciones epistémicas mínimas* que el Estado tiene la responsabilidad de proteger, pero que son estructurales y procedimentales, no sustantivas; conciernen a la *arquitectura* del espacio común, no a los *contenidos* que en él circulan. Cabe distinguir tres. Las de *acceso*: un acervo común de información de interés público (datos abiertos, transparencia, pluralismo de los medios); la libertad de prensa no es ajena al núcleo liberal, sino constitutiva de él. Las de *integridad de la validación*: la protección de la autonomía de las instituciones que producen y validan conocimiento (universidades, agencias científicas, organismos estadísticos, tribunales), de modo que su credibilidad descansa en su funcionamiento confiable y no en su subordinación al poder de turno; Kitcher (2011) ha mostrado, con su modelo de una *ciencia bien ordenada*, que

el conocimiento puede servir a una sociedad democrática sin quedar sometido al mercado ni al gobierno. Y las de *no manipulación*: la protección frente a técnicas que eluden la agencia racional, por ejemplo, la desinformación coordinada e inauténtica, la suplantación de identidad, la manipulación algorítmica opaca.

La distinción entre regular contenidos y regular arquitecturas es el eje normativo de la propuesta, y conviene darle forma de criterio. Propongo un *test de doble condición*: una medida de política epistémica es liberalmente admisible solo en la medida en que (i) pueda justificarse mediante razones públicas que prescindan de todo juicio sobre la verdad de las concepciones en disputa, condición de *neutralidad justificatoria*, y (ii) fortalezca, en lugar de sustituir, la agencia epistémica de los ciudadanos (condición de *agencia*). Una política liberal no debe pronunciarse sobre la verdad de las afirmaciones que circulan, pero sí puede intervenir sobre las condiciones bajo las cuales circulan. Exigir transparencia algorítmica, prohibir las redes coordinadas de cuentas falsas, obligar a revelar quién financia la publicidad política, garantizar la interoperabilidad: ninguna de estas medidas establece qué es verdadero, pero todas mejoran las condiciones bajo las cuales los ciudadanos pueden discernir la verdad por sí mismos. La analogía con el mercado es instructiva: el liberalismo no fija los precios, pero reprime el fraude y desmantela los monopolios; análogamente, una política epistémica liberal no fija las

creencias correctas, pero combate el fraude epistémico y disuelve los monopolios de la atención. Combatir una campaña de cuentas automatizadas que simulan un consenso inexistente pasa el test, porque la objeción no es que difundan falsedades, sino que engañan sobre su propia naturaleza; suprimir una opinión por considerarla falsa no lo pasa. El principio que orienta la frontera es claro: proteger las condiciones del juicio, no sustituir el juicio.

El test adquiere toda su urgencia ante el desafío que define la coyuntura tecnopolítica: la mediación algorítmica de la esfera pública. Lazar (2025) ha caracterizado el orden emergente como una *ciudad algorítmica*, es decir, la red de relaciones sociales mediadas por intermediarios algorítmicos; y ha mostrado que tales intermediarios, al constituir las relaciones que median, ejercen un *poder de gobierno* que reclama justificación o eliminación. Distingue tres preguntas que toda pretensión de poder debe responder: la del *qué* (la corrección sustantiva de las decisiones), la del *cómo* (la legitimidad procedimental) y la del *quién* (la autoridad propia para decidir). Su diagnóstico es que el debate sobre la “IA ética” se ha concentrado en la primera, en “alinear” los sistemas con valores correctos, desatendiendo las dos últimas, que son las propiamente políticas, y que las grandes empresas resuelven unilateral y opacamente.

La convergencia con el liberalismo epistémico es profunda. La propuesta perfeccionista, trasladada al entorno

digital, adopta hoy la forma de la *custodia algorítmica de la verdad*: la pretensión de que la moderación de contenidos, calibrada por clasificadores automáticos, depure el espacio público de la desinformación optimizando hacia resultados correctos. Esta tercera variante hereda todas las objeciones de las dos clásicas y agrega las suyas. Hereda la objeción falibilista (el clasificador, como el censor, debe presuponer un acceso a la verdad que la falibilidad invocada para justificarlo desmiente) y la objeción de la legitimidad, que un sistema acierte sobre el “qué” no responde al “quién” ni al “cómo”. Pero agrega un problema estructural que Douek (2021) ha diagnosticado: la moderación de contenidos a escala no es una secuencia de juicios individuales corregibles caso a caso, sino un sistema de administración masiva del discurso cuyas tasas de error son ineliminables y cuyas decisiones se sustraen, por su volumen y opacidad, a toda revisión significativa. El “atajo” algorítmico es, en la gramática de Lafont (2020), la forma más aguda de la deferencia ciega: el ciudadano no solo delega el juicio en un experto cuyas razones podría demandar, sino en un sistema cuyas razones son por diseño inescrutables.

De aquí se sigue una conclusión que el liberalismo epistémico fundamenta mejor que sus rivales. El problema normativo central de la gobernanza algorítmica de lo público no es primariamente que los sistemas se equivoquen sobre qué es verdadero, aunque se equivocan, sino que el poder de configurar el espacio común de razones se ha concentrado en

agentes que no satisfacen ninguna condición de legitimidad procedimental ni de autoridad propia. La respuesta liberal no consiste en exigir que las plataformas moderen “mejor” la verdad, lo que las consagraría como custodios, sino en someter su poder de gobierno a las dos condiciones del test: que sus intervenciones sobre la arquitectura del discurso sean públicamente justificables sin tomar partido sobre la verdad en disputa, y que fortalezcan en vez de suplantar la agencia de los usuarios. Transparencia de los sistemas de recomendación, contestabilidad de las decisiones que configuran la visibilidad, interoperabilidad que disperse el poder de los intermediarios, fricción deliberada en la viralización, desmantelamiento de las redes inauténticas: estas medidas operan sobre el *cómo* y el *quién* sin usurpar el juicio sobre el *qué*. El liberalismo epistémico ofrece, así, no una doctrina de la verdad digital, sino una teoría de la legitimidad del poder que la media.

La tradición neorrepública aporta aquí un complemento. La libertad como *no dominación* ilumina, con Pettit (1997), por qué la concentración del poder epistémico es un problema político de primer orden con independencia de cómo se ejerza en un momento dado. Una plataforma que controla el acceso de millones de ciudadanos a la información detenta un poder que, aun ejercido hoy con benevolencia, constituye una dominación potencial: los ciudadanos quedan a merced de decisiones que no controlan ni pueden contestar. Las condiciones epistémicas mínimas buscan, desde esta

óptica, menos producir creencias verdaderas que evitar que el espacio común quede sometido a un poder arbitrario, sea estatal, sea corporativo.

## **Objeciones, límites y conclusión**

La posición desarrollada enfrenta objeciones serias desde tres frentes. La primera, de cuño comunitarista, sostiene que el liberalismo es performativamente incoherente al apelar a virtudes cívicas: su antropología del individuo desvinculado erosiona los lazos densos de los que dependen esas virtudes. En la genealogía de MacIntyre (2004), las virtudes solo cobran sentido en prácticas y en una tradición narrativa que el individualismo moderno ha dismantelado; en la crítica de Sandel (1982), el “yo desvinculado” no logra dar cuenta de los lazos constitutivos de la identidad; en Taylor (1985), el atomismo presupone, sin poder justificarla, la matriz social que hace posible la agencia que exalta. La respuesta debe conceder lo verdadero de la objeción para refutar lo excesivo. Es cierto que las virtudes cívicas requieren prácticas que las cultiven; pero de ello no se sigue que exijan una comunidad de valores comprehensivos. La distinción rawlsiana entre un *consenso entrecruzado* y un consenso comprehensivo es decisiva: ciudadanos con concepciones del bien distintas pueden converger en virtudes cívicas desde sus respectivas doctrinas, cada uno por sus propias razones (Rawls, 2005b). La veracidad, la escucha y la responsabilidad pueden sostenerse por el creyente desde su fe, por el secular



desde su humanismo. Lo que el liberalismo niega no es la necesidad de virtudes, sino la de homogeneidad; y la objeción, al confundirlas, prueba demasiado: si las virtudes cívicas exigieran comunidad comprehensiva, ninguna sociedad moderna podría sostenerlas, lo que es empíricamente falso.

La segunda objeción procede del republicanismo y empuja en dirección contraria. Si la concentración del poder epistémico constituye dominación, el liberalismo sería demasiado tímido: la coherencia exigiría una intervención más enérgica, acaso la propiedad pública de las infraestructuras o una regulación sustantiva del discurso orientada a la igualdad de voz. Martí (2006) ha insistido en que la mera libertad negativa de expresión, en condiciones de desigualdad de recursos comunicativos, equivale a la libertad del más fuerte para imponer su voz. Concedo a esta objeción más que a la comunitarista, pues su preocupación es legítima y he intentado incorporarla en las condiciones epistémicas mínimas. Pero marco un límite: la intervención republicana sustantiva reintroduce, por la puerta de la no dominación, el perfeccionismo expulsado por la puerta de la verdad. Regular el discurso para garantizar la “igualdad efectiva de voz” exige juzgar qué voces merecen amplificación y cuáles son distorsión: juicios imposibles sin tomar partido sobre el contenido. La salida liberal distingue la *dispersión* del poder epistémico de su *captura* estatal: frente a la dominación corporativa, no transferir el poder al Estado, sino fragmentarlo (antimonopolio, interoperabilidad, pluralismo

de plataformas). Combatir la concentración no exige centralizar, sino dispersar.

La tercera objeción es la más urgente en el plano práctico. Ante amenazas graves y veloces: una pandemia en que la desinformación cuesta vidas, una injerencia electoral coordinada; los escrúpulos liberales parecen un lujo irresponsable. Tres consideraciones contienen su fuerza sin negar su parte de verdad. Primera: la lógica de la excepción tiende a expandirse; los poderes de emergencia para suprimir información, una vez normalizados, rara vez se repliegan y proveen a futuros gobiernos de herramientas de censura legitimadas por el precedente; quien diseña instituciones debe imaginarlas en manos de su peor adversario. Segunda: la eficacia de la supresión es dudosa, pues suprimir en un entorno de desconfianza confirma la narrativa conspirativa, convirtiendo al censurado en mártir. Tercera: existe un repertorio no perfeccionista, contradiscurso, etiquetado que añade contexto sin suprimir, fricción en la viralización, desmantelamiento de redes inauténticas, a menudo más eficaz que la supresión frontal. Reconozco casos límite, como la incitación directa e inminente a la violencia masiva; pero esos casos se gobiernan por los principios estrictos que el liberalismo ya posee para el daño directo, no por una nueva doctrina general de tutela de la verdad. La excepción genuina no funda la regla perfeccionista.

Respondidas las objeciones, la posición defendida se revela como un lugar estable, no un mero punto medio inestable. El *liberalismo epistémico* satisface cuatro condiciones que las posiciones rivales no logran reconciliar. Reconoce que la legitimidad democrática tiene una dimensión epistémica irreductible, pues sin cooperación epistémica no hay justificación mutua. Rechaza el perfeccionismo en todas sus formas (paternalista, epistocrática y algorítmico-tutelar) haciendo operar la protección de las condiciones epistémicas sobre la arquitectura del espacio público y no sobre sus contenidos. Sitúa en el centro la confianza pública, entendida como confiabilidad bien fundada. Y asigna un papel central a las virtudes cívicas, irreductibles a toda ingeniería institucional aunque tengan condiciones institucionales de cultivo. Responde, en suma, a una pregunta única: ¿cómo restaurar las condiciones epistémicas de la democracia sin destruir lo que la hace democrática? El perfeccionista responde produciendo buenas creencias, y acepta el costo autoritario; el *laissez-faire* responde no haciendo nada, y acepta la colonización del espacio público por el poder corporativo. El liberalismo epistémico responde construyendo y protegiendo las condiciones bajo las cuales ciudadanos libres, falibles y plurales pueden ejercer su juicio, confiar fundadamente y cooperar, sin que nadie (ni el Estado, ni los expertos, ni las plataformas) se arrogue la custodia de la verdad.

Arendt advirtió que el peligro totalitario no consiste primariamente en hacer creer la mentira, sino en destruir la capacidad de distinguir lo verdadero de lo falso: el sujeto ideal del dominio total es aquel para quien ya no existe la distinción entre hecho y ficción, ni entre lo verdadero y lo falso (Arendt, 1973). La crisis contemporánea no es totalitaria, pero participa de esa lógica de disolución. Su respuesta no fue jamás imponer una verdad oficial, que sería repetir la enfermedad, sino defender el espacio público plural como ámbito donde la pluralidad de perspectivas, lejos de disolver la realidad, la constituye intersubjetivamente. El liberalismo epistémico es heredero de esa intuición. Queda mucho por elaborar: la traducción institucional concreta de las condiciones epistémicas mínimas (qué exigir a las plataformas, cómo proteger la independencia de las instituciones de validación sin petrificarlas, cómo financiar un periodismo y una educación cívica a la altura del desafío) excede los límites de un trabajo de fundamentación normativa y reclama el concurso del derecho, la economía política y el diseño institucional. Pero la claridad sobre los principios es condición de la cordura en los detalles. La apuesta de este capítulo es que el liberalismo político y ético, lejos de ser un anacronismo superado por la era digital, conserva los recursos conceptuales para pensar una salida que no traicione ni la verdad ni la libertad. Que requiera ciudadanos virtuosos y no solo instituciones astutas es,

quizás, su lección más exigente y, a la vez, la más esperanzadora.

## Referencias

- Arendt, H. (1973). *The origins of totalitarianism* (Nueva ed.). Harcourt Brace Jovanovich. (Trabajo original publicado en 1951)
- Arendt, H. (2006). Truth and politics. En *Between past and future: Eight exercises in political thought* (pp. 223–259). Penguin Books.
- Brennan, J. (2016). *Against democracy*. Princeton University Press.
- Cassam, Q. (2019). *Vices of the mind: From the intellectual to the political*. Oxford University Press.
- Cortina, A. (1997). *Ciudadanos del mundo: Hacia una teoría de la ciudadanía*. Alianza Editorial.
- Cortina, A. (2017). *Aporofobia, el rechazo al pobre: Un desafío para la democracia*. Paidós.
- Douek, E. (2021). Governing online speech: From "posts-as-trumps" to proportionality and probability. *Columbia Law Review*, 121(3), 759–833.
- Estlund, D. M. (2008). *Democratic authority: A philosophical framework*. Princeton University Press.
- Forst, R. (2012). *The right to justification: Elements of a constructivist theory of justice*. Columbia University Press.

- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Goldman, A. I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63(1), 85–110. <https://doi.org/10.1111/j.1933-1592.2001.tb00093.x>
- Gutmann, A., & Thompson, D. (1996). *Democracy and disagreement*. Harvard University Press.
- Habermas, J. (1995). Reconciliation through the public use of reason: Remarks on John Rawls's political liberalism. *The Journal of Philosophy*, 92(3), 109–131. <https://doi.org/10.2307/2940842>
- Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Rehg, Trad.). MIT Press. (Trabajo original publicado en 1992)
- Han, B.-C. (2014). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder*. Herder.
- Han, B.-C. (2022). *Infocracia: La digitalización y la crisis de la democracia*. Taurus.
- Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.
- Lafont, C. (2020). *Democracy without shortcuts: A participatory conception of deliberative democracy*. Oxford University Press.

- Lazar, S. (2025). Governing the algorithmic city. *Philosophy & Public Affairs*, 53(2), 102–168.  
<https://doi.org/10.1111/papa.12279>
- Lynch, M. P. (2012). *In praise of reason: Why rationality matters for democracy*. MIT Press.
- Lynch, M. P. (2016). *The internet of us: Knowing more and understanding less in the age of big data*. Liveright.
- MacIntyre, A. (2004). *Tras la virtud* (2.<sup>a</sup> ed.). Crítica. (Trabajo original publicado en 1981)
- Martí, J. L. (2006). *La república deliberativa: Una teoría de la democracia*. Marcial Pons.
- Medina, J. (2013). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and resistant imaginations*. Oxford University Press.
- Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161.  
<https://doi.org/10.1017/epi.2018.32>
- O'Neill, O. (2002). *A question of trust: The BBC Reith Lectures 2002*. Cambridge University Press.
- Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Oxford University Press.
- Ranalli, C., & Lagewaard, T. (2022). Deep disagreement (Part 1): Theories of deep disagreement. *Philosophy Compass*, 17(12), e12886.  
<https://doi.org/10.1111/phc3.12886>
- Rauch, J. (2021). *The constitution of knowledge: A defense of truth*. Brookings Institution Press.

- Rawls, J. (1995). Reply to Habermas. *The Journal of Philosophy*, 92(3), 132–180.  
<https://doi.org/10.2307/2940843>
- Rawls, J. (2005a). *Political liberalism* (Ed. ampliada). Columbia University Press. (Trabajo original publicado en 1993)
- Rawls, J. (2005b). The idea of public reason revisited. En *Political liberalism* (Ed. ampliada, pp. 440–490). Columbia University Press. (Trabajo original publicado en 1997)
- Rosenfeld, S. (2018). *Democracy and truth: A short history*. University of Pennsylvania Press.
- Samaržija, H., & Cassam, Q. (Eds.). (2023). *The epistemology of democracy*. Routledge.
- Sandel, M. J. (1982). *Liberalism and the limits of justice*. Cambridge University Press.
- Saurette, P., & Gunster, S. (2011). Ears wide shut: Epistemological populism, argutainment and Canadian conservative talk radio. *Canadian Journal of Political Science*, 44(1), 195–218.  
<https://doi.org/10.1017/S0008423910001095>
- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- Taylor, C. (1985). Atomism. En *Philosophy and the human sciences: Philosophical papers 2* (pp. 187–210). Cambridge University Press.



- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Urbinati, N. (2014). *Democracy disfigured: Opinion, truth, and the people*. Harvard University Press.
- Vallier, K. (2019). *Must politics be war? Restoring our trust in the open society*. Oxford University Press.
- Vallier, K. (2021). *Trust in a polarized age*. Oxford University Press.
- Williams, B. (2002). *Truth and truthfulness: An essay in genealogy*. Princeton University Press.

# Transhumanismo, diseño cultural del futuro y el lugar de América Latina en el debate global

*Transhumanism, the Cultural Design of the Future, and Latin America's Place in the Global Debate*

**Constanza Rivano<sup>12</sup>**

## **Resumen**

Este capítulo examina el transhumanismo como una narrativa filosófica y tecnocultural del futuro humano. Primero, lo define como una filosofía de la anticipación que concibe al ser humano como una entidad abierta a la ampliación técnica de sus capacidades biológicas y cognitivas. Luego muestra que esta visión no permanece en el plano especulativo, sino que se articula en instituciones,

---

<sup>12</sup> Universidad de Chile. <https://orcid.org/0009-0009-6380-6152>

corporaciones tecnológicas y centros de investigación que participan activamente en la construcción material de ese futuro. A partir de ello, el texto analiza sus implicancias para América Latina, subrayando tensiones entre el imaginario transhumanista dominante y categorías como territorio, comunidad, memoria y vulnerabilidad. Finalmente, sostiene que es posible pensar un transhumanismo ético y democrático desde el Sur, siempre que se reoriente hacia la justicia estructural, la sostenibilidad ecológica, la pluralidad de futuros y la deliberación democrática.

**Palabras clave:** Transhumanismo; América Latina; Buen Vivir; futuro; filosofía de la tecnología.

## **Abstract**

This chapter examines transhumanism as a philosophical and technocultural narrative of the human future. First, it defines transhumanism as a philosophy of anticipation that understands the human being as open to the technical expansion of biological and cognitive capacities. It then argues that this vision does not remain at the speculative level, but is materially articulated through institutions, technology corporations, and research centers actively involved in shaping that future. On this basis, the chapter addresses its implications for Latin America, highlighting tensions between dominant transhumanist imaginaries and categories such as territory, community, memory, and

vulnerability. Finally, it argues that an ethical and democratic transhumanism can be imagined from the Global South, provided it is reoriented toward structural justice, ecological sustainability, plurality of futures, and democratic deliberation.

**Keywords:** Transhumanism; Latin America; Buen Vivir; future; philosophy of technology.

## **Introducción**

El transhumanismo se ha convertido en una de las narrativas más influyentes sobre el futuro humano. Sus promesas de longevidad radical, ampliación cognitiva, integración entre biología y técnica y eventual superación de ciertos límites corporales lo han situado en la intersección entre filosofía, innovación científica, economía tecnológica y geopolítica del conocimiento. Sin embargo, el transhumanismo no debe entenderse únicamente como un conjunto de hipótesis sobre el porvenir. También funciona como una narrativa cultural capaz de organizar expectativas colectivas, movilizar recursos materiales y redefinir qué formas de vida se consideran valiosas, deseables o legítimas. En este sentido, el problema filosófico no consiste solo en preguntarse si determinadas transformaciones serán posibles, sino en examinar qué idea de humanidad articulan, qué actores las impulsan y qué consecuencias producen en contextos históricamente desiguales.

Desde la filosofía de la tecnología, el futuro no se comprende simplemente como aquello que aún no ha ocurrido. Más profundamente, opera como una estructura de proyección desde la cual las sociedades orientan su acción en el presente. Las comunidades humanas no esperan pasivamente el porvenir: lo imaginan, lo anticipan y lo organizan mediante discursos, instituciones y dispositivos técnicos que configuran horizontes de posibilidad. Bajo esta lógica, el futuro se convierte en un campo de intervención y también en un espacio de disputa cultural. Las narrativas sobre lo que vendrá no solo describen posibilidades, sino que orientan inversiones, prioridades políticas y decisiones científicas. Por ello, el transhumanismo puede entenderse como una de las formulaciones más ambiciosas de esta imaginación tecnológica del futuro. Su fuerza no radica únicamente en las tecnologías que propone, sino en su capacidad para presentar una determinada visión del porvenir como deseable, plausible y, en ocasiones, inevitable.

Este capítulo se propone, en primer lugar, definir el transhumanismo como problema filosófico, situándolo en el cruce entre filosofía de la tecnología, estudios sobre anticipación y debates contemporáneos sobre el mejoramiento humano. En segundo lugar, busca mostrar que el transhumanismo no opera solo como doctrina abstracta, sino como diseño cultural del futuro, articulado a través de organizaciones, corporaciones tecnológicas, centros de investigación y discursos de innovación que convierten la

mejora humana en una realidad institucional y material. En tercer lugar, examina qué implica esta narrativa para América Latina, atendiendo a tensiones vinculadas con la dependencia tecnológica, la soberanía del porvenir y la vigencia de categorías como comunidad, territorio, memoria y vulnerabilidad. Finalmente, explora la posibilidad de una reformulación ética y democrática del transhumanismo desde el Sur, nutrida por el Buen Vivir, las epistemologías latinoamericanas y una concepción relacional de la vida humana.

Metodológicamente, el trabajo adopta un enfoque filosófico-crítico e interdisciplinario. Combina análisis conceptual del transhumanismo contemporáneo con una lectura político-cultural de sus principales actores, narrativas e implicancias. Este enfoque no pretende ofrecer una evaluación técnica exhaustiva del desarrollo científico actual ni una cartografía completa de todas las corrientes transhumanistas, sino interpretar el lugar que ocupa esta narrativa en la configuración contemporánea del futuro humano. Desde esta perspectiva, América Latina no aparece como una nota marginal añadida a un problema definido en otros lugares, sino como un espacio capaz de producir categorías filosóficas relevantes para pensar críticamente el presente tecnocientífico. Frente a una narrativa transhumanista que suele emerger desde ecosistemas del Norte global, la tradición crítica latinoamericana introduce preguntas distintas sobre justicia, relacionalidad, territorio y

vida compartida. Esa es, precisamente, la apuesta de estas páginas: mostrar que el debate sobre el futuro humano no está cerrado y que América Latina puede intervenir activamente en su resignificación.

## **Transhumanismo**

En la filosofía de la tecnología, el futuro no se entiende únicamente como aquello que todavía no ocurre. Más profundamente, funciona como una estructura de proyección desde la cual las sociedades orientan su acción en el presente. Las comunidades humanas no esperan pasivamente el porvenir: lo imaginan, lo anticipan y lo organizan mediante narrativas, instituciones y dispositivos técnicos que configuran horizontes de posibilidad. En este sentido, el futuro aparece como un espacio de disputa cultural donde se elaboran promesas de progreso, temores colectivos y proyectos civilizatorios. La tecnología participa activamente en esta dinámica porque modela expectativas, distribuye esperanzas y orienta decisiones políticas y económicas. La modernidad tecnocientífica puede comprenderse, por ello, como una cultura de la anticipación, en la que el porvenir se vuelve objeto de cálculo, planificación y diseño sistemático (Jasanoff, 2015). Este desplazamiento transforma la relación entre tiempo y acción: el futuro deja de percibirse como un destino incierto y comienza a configurarse como un campo de intervención. Bajo estas condiciones, las visiones sobre el porvenir adquieren una dimensión performativa, ya que no

solo describen lo que podría suceder, sino que también influyen en la manera en que las sociedades organizan sus prioridades, sus inversiones y sus imaginarios colectivos. Comprender esta lógica anticipatoria resulta fundamental para analizar el surgimiento del transhumanismo, pues esta corriente propone intervenir de manera deliberada en el diseño del futuro humano mediante el desarrollo de tecnologías emergentes.

El transhumanismo puede entenderse, en este contexto, como una de las formulaciones más ambiciosas de esta imaginación tecnológica del futuro. En términos generales, se presenta como una corriente filosófica y cultural que promueve el uso deliberado de la ciencia y la tecnología para ampliar las capacidades humanas y transformar la condición biológica actual. Desde esta perspectiva, el ser humano no aparece como una entidad fija o definitiva, sino como una etapa evolutiva abierta a procesos de modificación tecnológica (Bostrom, 2005). Bajo esta mirada, la enfermedad, el envejecimiento o la fragilidad cognitiva dejan de interpretarse como rasgos inevitables de la existencia y pasan a considerarse problemas susceptibles de intervención científica. El transhumanismo retoma así una tradición intelectual vinculada al ideal ilustrado de perfectibilidad humana, aunque desplaza el centro de la transformación hacia la ingeniería biológica, la inteligencia artificial y la convergencia tecnológica. Dentro de este horizonte, uno de sus conceptos más influyentes ha sido el de singularidad



tecnológica, entendido como un escenario en el que la inteligencia artificial alcanzaría niveles superiores a la inteligencia humana y produciría transformaciones profundas en la organización del conocimiento y de la vida social (Kurzweil, 2006). La fuerza cultural de esta narrativa reside en que ofrece una promesa de trascendencia intramundana basada en la innovación tecnológica. De este modo, el transhumanismo no solo describe un futuro posible, sino que contribuye a redefinir lo que las sociedades contemporáneas consideran deseable para el porvenir humano.

A pesar de la fuerza de estas promesas, el transhumanismo también plantea interrogantes éticos y políticos que requieren un examen riguroso. La ampliación de capacidades humanas mediante innovación tecnológica ocurre dentro de contextos marcados por desigualdades económicas, asimetrías geopolíticas y diferencias profundas en el acceso a recursos científicos. La pregunta por quién podrá beneficiarse de estas tecnologías adquiere entonces una relevancia central. La expansión de capacidades cognitivas o biológicas podría generar escenarios de acceso diferencial a tecnologías de mejora, introduciendo nuevas formas de estratificación social basadas en capacidades tecnológicamente amplificadas. Desde la ética de la tecnología, esta cuestión no puede evaluarse únicamente en términos de eficiencia tecnológica o innovación científica, porque exige considerar también las condiciones sociales que

determinan la distribución de los beneficios del progreso (Vallor, 2016). En este contexto, el transhumanismo aparece vinculado a debates más amplios sobre justicia tecnológica, gobernanza científica y responsabilidad colectiva. La promesa de rediseñar el cuerpo, prolongar la vida o ampliar la inteligencia abre horizontes inéditos para la historia humana, pero también obliga a examinar sus implicaciones culturales y políticas. En consecuencia, el transhumanismo emerge como un campo de discusión filosófica en el que convergen preguntas sobre progreso científico, equidad social y sentido de la vida humana en una época marcada por transformaciones tecnológicas aceleradas.

Estas tensiones se vuelven todavía más visibles cuando el transhumanismo se analiza a escala global. Las narrativas que articulan este proyecto emergen principalmente en contextos tecnológicos del Norte global, donde confluyen centros de innovación científica, grandes inversiones en investigación y ecosistemas empresariales de alta complejidad. En esos espacios se produce una visión del futuro humano caracterizada por la expansión de capacidades, la integración entre biología y tecnología y la aceleración del desarrollo científico. Sin embargo, este horizonte no coincide necesariamente con las prioridades de otras regiones del mundo. En contextos atravesados por desigualdades estructurales, conflictos territoriales o crisis ambientales, las promesas de una humanidad tecnológicamente ampliada adquieren significados

diferentes. Desde perspectivas críticas del Sur global se ha señalado que el futuro no puede concebirse como una trayectoria única y universal, ya que las sociedades elaboran sus horizontes de desarrollo desde experiencias históricas y contextos culturales diversos (Santos, 2010). Bajo esta mirada, el transhumanismo aparece como una forma específica de imaginar el porvenir humano, vinculada a determinados centros de poder tecnológico y económico. Su ambición es indudablemente notable: propone intervenir en los procesos evolutivos mediante el conocimiento científico y transformar las condiciones biológicas de la existencia. Al mismo tiempo, esa ambición requiere ser examinada con cautela. Analizar el transhumanismo implica preguntarse hasta qué punto las sociedades desean orientar recursos políticos, científicos y culturales hacia la realización de este diseño tecnológico del futuro.

## **Diseño cultural**

Si el transhumanismo propone una filosofía del futuro humano, entonces la primera pregunta que debe formularse no es únicamente qué tecnologías harán posible ese futuro, sino quiénes lo imaginan y desde qué lugares se proyecta. En las sociedades contemporáneas, el porvenir ha dejado de aparecer solo como una continuidad temporal abierta y se ha convertido en un espacio cultural activamente diseñado mediante discursos, instituciones, dispositivos técnicos y marcos de interpretación que orientan las expectativas

colectivas. Imaginar el futuro implica, por ello, una forma de poder, en la medida en que esas narrativas influyen sobre la organización del conocimiento, la dirección de la innovación tecnológica y las prioridades de la acción política en el presente (Jasanoff, 2015). Desde esta perspectiva, el transhumanismo no debe entenderse únicamente como una corriente filosófica interesada en la mejora humana, sino también como una narrativa cultural capaz de reorganizar el horizonte de expectativas de la humanidad. La imagen de una especie ampliada por la tecnología, optimizada mediante algoritmos y liberada de diversas limitaciones biológicas no permanece en el plano abstracto, sino que orienta proyectos científicos, inversiones tecnológicas y agendas institucionales. Bajo estas condiciones, el diseño cultural del futuro adquiere una dimensión política profunda, ya que no solo define qué innovaciones se consideran deseables, sino también qué formas de humanidad comienzan a ser presentadas como legítimas y valiosas para las próximas décadas.

Esta lógica se intensifica cuando el debate se traslada al contexto latinoamericano, donde la cuestión del futuro tecnológico no puede separarse de trayectorias históricas marcadas por desigualdades estructurales, dependencia científica y experiencias de colonización. Gran parte de las narrativas globales sobre el porvenir provienen de centros de poder científico y económico ajenos a la región, y circulan muchas veces como si representaran aspiraciones universales

de la humanidad. Desde una perspectiva crítica situada en el Sur, esta universalización requiere ser examinada con cautela, pues las tradiciones intelectuales latinoamericanas han tendido a pensar lo humano de manera relacional, histórica y territorial, en tensión con la imagen del individuo optimizable que domina buena parte del imaginario transhumanista (Santos, 2010). Esto conduce a una pregunta todavía más profunda sobre la universalidad de lo humano. Si el transhumanismo define al sujeto a partir de la autonomía, la mejora tecnológica y el rediseño de sí, diversas corrientes latinoamericanas insisten, en cambio, en la centralidad de la memoria histórica, la vulnerabilidad compartida y la interdependencia social para comprender la condición humana (Mignolo, 2005). Esta tensión se vuelve aún más urgente cuando se considera que América Latina participa de manera desigual en la producción de tecnologías emergentes y, en muchos casos, debe consumir o adaptar desarrollos realizados en otros centros de poder. Por eso, la discusión sobre el transhumanismo no remite solo a la expansión de capacidades humanas, sino también a problemas de soberanía tecnológica, regulación ética y justicia distributiva. En un escenario donde algunos imaginan la trascendencia de la figura humana mediante innovación científica, la pregunta decisiva sigue abierta: ¿quiénes participarán realmente de ese futuro?

## **Mapa de actores transhumanistas**

Antes de profundizar en el análisis del transhumanismo desde una perspectiva latinoamericana, conviene dimensionar la magnitud institucional, económica y científica que hoy sostiene este movimiento a escala global. El transhumanismo ya no es solo una corriente filosófica que imagina futuros posibles para la humanidad. En las últimas décadas se ha consolidado como un ecosistema de investigación, innovación tecnológica y producción cultural que involucra universidades, centros de pensamiento, organizaciones políticas, corporaciones tecnológicas y laboratorios de biotecnología. Estos actores no se limitan a especular sobre el porvenir, sino que participan activamente en la construcción de tecnologías destinadas a ampliar capacidades humanas, prolongar la vida biológica y reconfigurar la relación entre cuerpo, mente y técnica. En este sentido, el transhumanismo opera como un campo institucional donde convergen intereses científicos, financieros y culturales orientados a redefinir el horizonte humano.

Entre las organizaciones más influyentes destaca Humanity+, fundada a fines de la década de 1990 por Nick Bostrom y David Pearce. Su relevancia radica en haber contribuido a consolidar los fundamentos filosóficos y éticos del movimiento, promoviendo el uso de tecnologías emergentes para expandir las capacidades humanas y

mejorar las condiciones de vida. A través de conferencias, publicaciones y redes de investigación, Humanity+ ha articulado un lenguaje común para debates sobre longevidad, inteligencia artificial, aumento cognitivo y transferencia digital de la mente, proyectando estas ideas más allá del ámbito académico y otorgándoles visibilidad cultural (Bostrom, 2005).

Junto a este plano doctrinal, el transhumanismo también se ha institucionalizado en espacios orientados a la innovación y al diseño estratégico del futuro. Un caso emblemático es Singularity University, fundada por Ray Kurzweil y Peter Diamandis en Silicon Valley. Aunque se presenta como una institución educativa, funciona más bien como un think tank dedicado a promover tecnologías exponenciales y su aplicación a distintos ámbitos de la vida social. Sus programas reúnen a empresarios, tecnólogos y tomadores de decisión, convirtiéndose en una plataforma clave para traducir la narrativa de la singularidad en proyectos concretos de innovación y emprendimiento (Morozov, 2013).

A este entramado se suman las grandes corporaciones tecnológicas, que cumplen un rol central en la materialización del imaginario transhumanista. Empresas como Google, Microsoft, Meta, Apple y OpenAI investigan áreas decisivas como inteligencia artificial, neurotecnología, realidad aumentada y análisis biomédico de datos. Aunque no se

definan explícitamente como transhumanistas, sus laboratorios y proyectos —como DeepMind o Neuralink— trabajan en tecnologías que apuntan directamente a la ampliación cognitiva, la interfaz cerebro-máquina o la integración entre procesos biológicos y sistemas artificiales. La escala de inversión involucrada muestra que el transhumanismo ha dejado de ser una especulación filosófica marginal para convertirse en un campo estratégico de innovación científica y económica (Bratton, 2021).

Otro eje fundamental de este mapa de actores está constituido por las instituciones dedicadas a la longevidad y la biotecnología radical. Organizaciones como la SENS Research Foundation y diversos Longevity Institutes impulsan investigaciones orientadas a intervenir en los procesos biológicos del envejecimiento, mientras empresas biotecnológicas y startups médicas desarrollan terapias celulares, ingeniería genética y tratamientos de rejuvenecimiento. Las proyecciones de mercado muestran que este sector seguirá expandiéndose sostenidamente hacia 2030, alcanzando cifras de inversión de gran magnitud (Allied Market Research, 2022). En conjunto, este panorama revela que el transhumanismo se ha consolidado como un proyecto global que articula pensamiento filosófico, investigación científica e inversión económica en torno a la transformación del ser humano. Comprender quiénes lo impulsan y con qué recursos resulta fundamental para evaluar sus implicancias filosóficas y políticas, especialmente



en regiones como América Latina, que deberán situar críticamente su posición frente a un proyecto tecnológico global que avanza con rapidez hacia la próxima década.

## **Latinoamérica frente al futuro Ttranshumanista**

Con la magnitud institucional, económica y científica que hoy moviliza el transhumanismo, resulta cada vez más difícil sostener que América Latina pueda permanecer al margen de esta discusión. Aunque los desarrollos más avanzados en inteligencia artificial, biotecnología y neurotecnología siguen concentrándose en centros de innovación del Norte global, sus efectos políticos y culturales alcanzan de forma creciente a las sociedades latinoamericanas. La región ya participa indirectamente en estas transformaciones mediante procesos de adopción tecnológica, cooperación científica, debates regulatorios y políticas públicas ligadas a la digitalización. Sin embargo, el problema no consiste solo en determinar si América Latina logrará integrarse a la nueva ola tecnológica global. La pregunta más profunda se refiere a la capacidad de la región para definir su propio horizonte de futuro.

Pensar el transhumanismo desde América Latina implica preguntarse qué tipo de sociedad se quiere construir y bajo qué principios culturales debe orientarse la relación entre tecnología y vida humana. Desde la filosofía de la liberación, el futuro latinoamericano no puede reducirse a

una modernización imitativa que reproduzca modelos surgidos en otros contextos históricos, porque la región posee la capacidad de producir conceptos propios y de formular alternativas civilizatorias distintas de aquellas que emergieron en la modernidad euroatlántica (Dussel, 2004). En este sentido, el debate sobre el transhumanismo abre una cuestión decisiva: si América Latina desea simplemente adaptarse a un futuro diseñado por otros o si pretende participar activamente en la construcción de un horizonte histórico propio. Esta diferencia es crucial, porque transforma el problema tecnológico en una interrogación filosófica sobre autonomía cultural, imaginación política y soberanía del porvenir.

Esta interrogante adquiere todavía mayor densidad cuando se examina la relación histórica entre tecnología y poder en el continente. A lo largo de la modernidad, América Latina ha experimentado la incorporación de tecnologías en marcos económicos y políticos marcados por la dependencia, la extracción de recursos y la subordinación al mercado global. La tecnología ha funcionado muchas veces como un instrumento integrado a lógicas de valorización externa antes que como un medio orientado a la emancipación regional. Por ello, resulta necesario distinguir entre el potencial transformador de la tecnología y las formas históricas de poder que determinan su uso. Bolívar Echeverría observó que la tecnología no debe ser identificada automáticamente con la dominación, ya que la modernidad contenía una promesa

emancipadora vinculada al despliegue de las capacidades humanas, pero esa promesa quedó subsumida bajo la racionalidad capitalista que organiza la técnica en función de la acumulación económica (Echeverría, 1998). Desde esta perspectiva, el transhumanismo aparece como una expresión contemporánea de esa tensión. Por un lado, promete ampliar las capacidades humanas mediante tecnologías capaces de mejorar la salud, la cognición y la longevidad. Por otro lado, estas promesas emergen dentro de un sistema global profundamente desigual, donde el acceso a la innovación se distribuye de manera extremadamente asimétrica. Para América Latina, esta situación impone un desafío político inmediato: evaluar hasta qué punto la expansión tecnológica contemporánea reproduce estructuras de dependencia que ya han marcado la historia de la región. La pregunta por la mejora humana no puede desvincularse, entonces, de la pregunta por quién controla la infraestructura tecnológica, quién define sus fines y bajo qué condiciones circulan sus beneficios.

La discusión se profundiza aún más cuando se considera la identidad cultural latinoamericana. La región no puede comprenderse solo a través de categorías abstractas de progreso tecnológico o modernización económica, porque su historia está atravesada por colonización, mestizaje, resistencia cultural y búsqueda de reconocimiento político. Rodolfo Kusch mostró que América Latina no se define únicamente por su incorporación incompleta a la

modernidad, sino por la persistencia de formas de vida que no encajan plenamente dentro de la racionalidad moderna occidental. Su reflexión sobre el “estar” frente al “ser” revela una experiencia histórica arraigada en el territorio, la comunidad y la memoria colectiva (Kusch, 2000). Esta perspectiva resulta especialmente significativa cuando se analiza el imaginario transhumanista. Mientras el transhumanismo imagina el futuro como una expansión tecnológica orientada a superar los límites biológicos de la especie humana, gran parte del pensamiento latinoamericano ha desarrollado una sensibilidad filosófica que enfatiza la coexistencia con el territorio, la memoria histórica y la vulnerabilidad compartida. De este modo, la cuestión tecnológica se convierte también en una pregunta cultural sobre el tipo de humanidad que una sociedad desea afirmar.

Este problema se vincula estrechamente con la noción de colonialidad del poder desarrollada por Aníbal Quijano, quien mostró que la modernidad no solo organizó el mundo mediante jerarquías económicas y políticas, sino también mediante jerarquías epistémicas que definieron qué formas de conocimiento eran consideradas válidas (Quijano, 1992). El discurso transhumanista participa de esta dinámica cuando universaliza una visión del futuro centrada en la mejora tecnológica del individuo, la expansión cognitiva y la prolongación radical de la vida. Desde América Latina, esta universalización exige una crítica filosófica rigurosa, porque las prioridades históricas de la región han estado marcadas

por desigualdad social, defensa del territorio, acceso a derechos básicos y reconocimiento cultural. En este contexto, la promesa de trascender los límites biológicos mediante tecnología aparece como una aspiración profundamente ambivalente, cuya pertinencia histórica no puede darse por supuesta.

Es precisamente en este punto donde el Buen Vivir adquiere una relevancia filosófica decisiva. En su formulación kichwa como *Sumak Kawsay* y en su variante aymara como *Suma Qamaña*, el Buen Vivir no constituye simplemente una política social alternativa ni un correctivo más humano del desarrollo económico. Se trata de una crítica profunda a la matriz civilizatoria de la modernidad occidental, en especial a su versión capitalista, extractivista y antropocéntrica. El Buen Vivir propone una comprensión relacional de la existencia humana, donde la vida buena no se define por la acumulación ilimitada ni por la maximización del rendimiento individual, sino por el equilibrio entre comunidad, naturaleza y sostenibilidad de la vida colectiva (Acosta, 2010). Esta concepción introduce una diferencia filosófica fundamental frente al imaginario transhumanista. Mientras el transhumanismo privilegia la ampliación tecnológica de las capacidades individuales, el Buen Vivir propone una ética de la interdependencia que sitúa a los seres humanos dentro de una red más amplia de relaciones con el territorio, la memoria y los ciclos naturales. Arturo Escobar ha desarrollado esta dimensión ontológica al interpretar el Buen Vivir como parte

de una concepción relacional del mundo que desafía la separación moderna entre naturaleza y sociedad, y que permite pensar un pluriverso de formas de vida en lugar de un único modelo universal de desarrollo (Escobar, 2018).

Desde esta perspectiva, el debate sobre el futuro tecnológico deja de girar exclusivamente en torno a la fascinación por la innovación y se transforma en una reflexión sobre qué mundos producen determinadas tecnologías y bajo qué criterios esos mundos pueden considerarse habitables. La tensión entre el imaginario transhumanista y el horizonte del Buen Vivir revela, por tanto, que la discusión sobre el futuro tecnológico no puede separarse de las preguntas por identidad, soberanía y justicia. América Latina se encuentra en una posición histórica donde debe decidir cómo relacionarse con las transformaciones tecnológicas de las próximas décadas: mantenerse al margen podría profundizar la dependencia, pero adoptar acríticamente las narrativas dominantes podría reforzar modelos incompatibles con sus realidades sociales y ambientales. En última instancia, el debate abre una cuestión filosófica de gran alcance: cómo imaginar el futuro sin reproducir las jerarquías coloniales que han estructurado la historia moderna. Frente a la promesa transhumanista de trascender los límites humanos mediante innovación tecnológica, el Buen Vivir recuerda que el desafío más urgente del siglo XXI podría no consistir en ir más allá de lo humano, sino en aprender a vivir mejor dentro de los límites que hacen posible la vida compartida.

## **¿Un transhumanismo democrático y ético? Una posibilidad desde el Sur**

Llegados a este punto, la pregunta ya no puede formularse únicamente en términos de adhesión o rechazo frente al transhumanismo. Después de examinar su arquitectura conceptual, sus actores globales y las tensiones que genera en América Latina, el problema adquiere otra densidad: ¿Tiene realmente futuro Latinoamérica dentro del proyecto transhumanista?, y, más aún, ¿Puede ese proyecto ser reformulado de manera ética y democrática desde una perspectiva del Sur? La respuesta no parece residir en una simple incorporación periférica a un diseño tecnológico elaborado en otros centros de poder, ni tampoco en una negativa nostálgica frente a toda innovación técnica. Lo decisivo consiste en reconocer que el transhumanismo, tal como hoy se presenta, carga con una fuerte impronta elitista, tecnocrática y universalista, aunque no por ello esté clausurado a otras posibilidades históricas. América Latina no ocupa una posición central en la formulación de esta narrativa del futuro, pero precisamente por ello puede ofrecer una interpelación crítica capaz de alterar su sentido. Desde ahí se abre una hipótesis fértil: quizá la cuestión no consista simplemente en averiguar hasta dónde la tecnología podrá superar ciertos límites humanos, sino en decidir qué tipo de convivencia hará posible y qué forma de mundo ayudará a sostener. En ese desplazamiento, el proyecto transhumanista dejaría de aparecer como una carrera solitaria hacia la

perfección y podría comenzar a pensarse como una práctica de transformación situada, compartida y responsable.

Una de las primeras condiciones para que esta relectura sea posible consiste en revisar el ideal de mejora que el transhumanismo ha heredado de la modernidad occidental. En muchas de sus versiones dominantes, mejorar significa ampliar potencia, prolongar la vida, incrementar rendimiento cognitivo y reducir fragilidad. Este imaginario organiza la tecnología como una promesa de emancipación respecto del límite, el dolor, la dependencia y la finitud. Sin embargo, la tradición latinoamericana del Buen Vivir introduce una objeción decisiva a esta lógica. Como vimos, allí la vida buena no se define por la acumulación de capacidades ni por la expansión indefinida de opciones individuales, sino por la armonía entre sujetos, comunidad, naturaleza y formas de habitar el mundo (Acosta, 2013). Esta diferencia altera el corazón normativo del debate. Obliga a reconocer que no toda ampliación tecnológica equivale por sí misma a una mejora ética. Una sociedad puede aumentar sus capacidades y, al mismo tiempo, profundizar el aislamiento, la competencia y la devastación ambiental. Una tecnología puede optimizar funciones biológicas y, a la vez, erosionar las tramas de reciprocidad que sostienen una vida común. Desde esta perspectiva, un transhumanismo ético no podría organizarse solo alrededor de la pregunta por lo que los cuerpos o las mentes pueden llegar a hacer, sino también en torno a cómo esas transformaciones afectan el modo en que convivimos,



cuidamos y compartimos mundo. El aporte latinoamericano consiste precisamente en recordar que la tecnología no debe orientarse únicamente al perfeccionamiento de individuos, sino a la habitabilidad de la vida colectiva.

Este punto permite imaginar una mutación más profunda del horizonte transhumanista. Si se lo piensa desde el Sur, el objetivo de una tecnología que excede lo humano no tendría por qué reducirse a superar capacidades y habilidades en un sentido competitivo. Podría orientarse, más bien, a promover la adaptabilidad entre comunidades y territorios, fortaleciendo capacidades de cohabitación más justas y sostenibles. Esta hipótesis implica un cambio de eje considerable. La inteligencia tecnológica dejaría de concebirse como un instrumento para imponerse sobre otros cuerpos, otros ecosistemas y otras culturas, y comenzaría a entenderse como una mediación para vivir con otros de manera más respetuosa. En esta clave, las tecnologías del futuro podrían contribuir a enfrentar crisis ambientales, desigualdades sanitarias, desplazamientos poblacionales o vulnerabilidades estructurales desde una lógica de cooperación y no de supremacía. La innovación ya no se mediría únicamente por su capacidad de producir individuos más eficientes, sino por su potencia para hacer más habitable la coexistencia entre diferencias humanas, territoriales y ecológicas. Desde el punto de vista filosófico, esto exige pasar de una ontología de la superación a una ontología de la relación. El valor de la tecnología no residiría solo en su poder

de transformar organismos, sino en su capacidad de reconfigurar vínculos sin destruirlos. América Latina puede ofrecer aquí una intuición decisiva: la verdadera transformación no consiste necesariamente en escapar de la condición situada, sino en aprender a habitarla con más inteligencia, más cuidado y más justicia. Un transhumanismo democrático tendría que asumir esa lección si aspira a ser algo más que una teología secular del rendimiento.

A esta reorientación ética se suma un segundo aporte fundamental: la crítica latinoamericana de la colonialidad del saber y del poder. El transhumanismo suele presentarse como una narrativa universal sobre el futuro de la especie, como si sus imágenes del cuerpo, de la inteligencia y de la vida deseable expresaran una aspiración común de toda la humanidad. Sin embargo, desde la filosofía decolonial sabemos que ninguna idea del futuro es neutral; toda proyección nace desde una localización histórica, geopolítica y epistémica determinada (Mignolo, 2005). El transhumanismo dominante surge en ecosistemas tecnocientíficos muy específicos, marcados por capital intensivo, infraestructura avanzada, individualismo liberal y una confianza casi dogmática en la innovación como motor civilizatorio. Presentarlo como destino general del ser humano invisibiliza esa localización y desautoriza otras formas posibles de imaginar el porvenir. La intervención latinoamericana resulta valiosa porque obliga a desnaturalizar esa universalidad. Al hacerlo, no pretende

clausurar el debate transhumanista, sino abrirlo a su propia historicidad. Un transhumanismo ético y democrático tendría que reconocer que su imagen del humano mejorado no representa automáticamente a toda la humanidad, y que existen otras experiencias históricas, otras cosmologías y otros deseos colectivos que no se dejan reducir al mismo modelo de progreso. Esta corrección no debilita el proyecto; lo vuelve filosóficamente más serio, porque lo libera de la pretensión de hablar en nombre del todo. Desde el Sur, el futuro deja de ser un mandato unívoco y se convierte en una conversación conflictiva entre mundos, memorias y aspiraciones distintas.

Esa apertura exige también incorporar una conciencia más aguda de las asimetrías políticas de la tecnociencia. Con demasiada frecuencia, el transhumanismo piensa la técnica en clave de posibilidad, como si bastara con preguntarse qué innovaciones permitirán ampliar la vida, la inteligencia o la sensibilidad humanas. América Latina introduce otra pregunta, igualmente decisiva: ¿Quién produce esas tecnologías, quién controla sus infraestructuras y bajo qué relaciones de poder circulan? En una región marcada por brechas en salud, educación, conectividad y soberanía digital, la promesa de mejora biotecnológica no puede analizarse al margen de la dependencia estructural. No es lo mismo imaginar la ampliación humana desde un centro global de innovación que hacerlo desde territorios donde el acceso a derechos básicos sigue siendo desigual. Esta diferencia

modifica el sentido mismo del debate. Un transhumanismo democrático no podría limitarse a garantizar el derecho abstracto a modificar el cuerpo o prolongar la vida; tendría que pensar las condiciones sociales, económicas y geopolíticas que hacen posible o imposible ese derecho. En este punto, la filosofía latinoamericana desplaza el problema desde una ontología de la optimización hacia una ontología política de la tecnociencia. La cuestión ya no gira solo en torno a cuánto puede transformarse el cuerpo humano, sino alrededor de qué estructuras de poder organizan esa transformación. Tal desplazamiento resulta indispensable si se quiere evitar que la mejora técnica se convierta en una nueva forma de segregación biológica y civilizatoria.

Un cuarto aporte decisivo proviene de una comprensión más densa de la corporeidad. Aunque el transhumanismo habla constantemente del cuerpo, con frecuencia lo hace bajo una mirada funcional: cuerpo como plataforma actualizable, interfaz corregible, soporte mejorable. Este enfoque pierde de vista que el cuerpo humano no es solo un conjunto de capacidades cuantificables, sino también una forma de exposición, de memoria y de inscripción histórica. La filosofía latinoamericana contemporánea, especialmente cuando dialoga con feminismos, fenomenología y pensamiento crítico, permite recuperar esa complejidad. El cuerpo aparece entonces atravesado por clase, raza, género, territorio, violencia y cuidado; no es un objeto neutro disponible para la ingeniería,

sino una experiencia vivida situada en tramas de poder y vulnerabilidad. Esta relectura importa mucho para un transhumanismo que aspire a ser ético. Le recordaría que transformar técnicamente el cuerpo no equivale simplemente a optimizar un sistema biológico, porque toda intervención se inscribe sobre vidas concretas, sobre historias marcadas y sobre desigualdades acumuladas. En ese marco, la vulnerabilidad ya no puede ser leída solo como déficit técnico. También es condición de nuestra coexistencia. Nos constituye como seres que dependen de otros, que requieren mundo común, que no pueden reducir su dignidad a la suma de sus funciones. Desde América Latina, esta concepción del cuerpo obliga a moderar la fantasía de autosuficiencia que muchas veces acompaña al ideal transhumanista y a pensar la transformación técnica desde la experiencia concreta de la fragilidad compartida.

La reflexión sobre la vulnerabilidad conduce directamente a la pregunta por la democracia. Un transhumanismo democrático sería aquel que reconoce que toda modificación del cuerpo, de la mente o de la conciencia tiene consecuencias colectivas y, por tanto, no puede decidirse únicamente en laboratorios privados, corporaciones tecnológicas o enclaves de expertos. La democracia, en este contexto, no consiste solo en distribuir beneficios al final del proceso, sino en abrir desde el inicio los lenguajes, las prioridades y los criterios de decisión sobre aquello que una sociedad considera valioso transformar. Esto implica

procesos deliberativos amplios, inteligibles y culturalmente situados, donde participen comunidades, pueblos originarios, organizaciones sociales, mujeres, personas con discapacidad, sujetos neurodivergentes y todas aquellas formas de vida históricamente marginadas del diseño de lo humano. En este sentido, una ética de las tecnologías emergentes necesita fundarse en virtudes más amplias que la eficiencia y la innovación. Requiere justicia, compasión, prudencia y sabiduría práctica (Vallor, 2016). América Latina puede aportar mucho a este horizonte porque su historia política ha estado atravesada por luchas por reconocimiento, derechos humanos, memoria y participación. Estas experiencias enseñan que la dignidad no puede definirse desde arriba ni reducirse a un parámetro técnico. Si el transhumanismo quiere volverse democrático, deberá aceptar que el futuro tecnológico no pertenece solo a quienes financian la investigación, sino a quienes habitarán sus consecuencias. Esa ampliación del demos es, quizá, una de las transformaciones más urgentes que el proyecto necesita.

## **Viabilidad**

Ahora bien, para que un transhumanismo nutrido por la filosofía latinoamericana fuera realmente viable, no bastaría con adoptar un discurso más inclusivo o una retórica crítica de la desigualdad. Necesitaría un modelo institucional distinto del que hoy suele acompañar a las promesas transhumanistas dominantes. Es decir, no podría organizarse

principalmente en torno al mercado, la innovación corporativa y la elección individual, sino alrededor de una arquitectura pública, democrática, regional y éticamente orientada al bien común. El modelo más adecuado sería algo así como un modelo público-deliberativo de innovación tecnocientífica, capaz de articular soberanía tecnológica, justicia social y deliberación democrática. Este modelo requeriría, en primer lugar, una inversión estatal y regional robusta en ciencia y tecnología, no para competir miméticamente con Silicon Valley, sino para desarrollar capacidades soberanas en ámbitos estratégicos como inteligencia artificial, neurotecnología, salud digital, biotecnología y regulación de datos. En segundo lugar, exigiría una gobernanza democrática de las tecnologías emergentes, donde las decisiones sobre automatización cognitiva o modificación biológica no queden solo en manos de expertos, empresas o capitales privados, sino que incluyan deliberación pública, participación interdisciplinaria y representación social. En tercer lugar, supondría una institucionalidad regional latinoamericana, porque ningún país de la región posee por sí solo la escala suficiente para sostener soberanía tecnocientífica robusta. En cuarto lugar, requeriría una articulación estrecha entre universidades, investigación científica, salud pública y política social, de modo que la innovación no se diseñe solo como ventaja competitiva, sino como infraestructura de justicia. Finalmente, precisaría una ética regulatoria preventiva, capaz

de intervenir antes de que las tecnologías produzcan desigualdades irreversibles (UNESCO, 2021).

Un modelo de esta naturaleza implicaría que el mejoramiento humano, en caso de impulsarse, no sería tratado como un bien de consumo *premium* ni como un privilegio de mercado, sino como una cuestión de orientación civilizatoria. Por eso, el criterio de viabilidad institucional no es solo técnico o financiero; es ante todo político. América Latina necesitaría instituciones capaces de preguntar no solo qué innovación es posible, sino cuál merece ser promovida, bajo qué fines, con qué límites y con qué mecanismos de distribución. Esta diferencia es decisiva, porque desplaza el eje del debate desde la fascinación por la capacidad técnica hacia la responsabilidad pública sobre los efectos de esa capacidad. En lugar de asumir que toda innovación valiosa debe abrirse camino por la lógica del emprendimiento y del capital privado, un transhumanismo desde el Sur exigiría que la investigación se encuentre subordinada a fines sociales explícitos. La pregunta no sería únicamente cómo acelerar el progreso, sino para quién, para qué y bajo qué condiciones. Aquí aparece un contraste importante con la arquitectura dominante de la innovación transhumanista, hoy muy ligada a consorcios corporativos, filantropía tecnológica y competencia geopolítica. Frente a esa configuración, la apuesta latinoamericana tendría que insistir en que el futuro humano no puede depender exclusivamente de quienes poseen las infraestructuras del cálculo, la inversión y los



datos. Si la técnica ha de rediseñar algo tan decisivo como el cuerpo, la cognición o la longevidad, entonces debe quedar sometida a una institucionalidad capaz de responder ante la sociedad y no solo ante el mercado.

Por consiguiente, para diferenciarse realmente del tecnoprogresismo liberal, este modelo institucional necesita también criterios normativos propios. Ese es el punto decisivo. El tecnoprogresismo liberal acepta la ampliación tecnológica de capacidades, pero intenta corregir sus excesos mediante acceso más amplio, regulación razonable e inclusión distributiva. Sigue siendo, sin embargo, liberal en su núcleo: supone que la mejora es en principio deseable, que el individuo debe conservar un alto grado de soberanía sobre su autotransformación y que el problema principal consiste en distribuir mejor los beneficios del progreso. Un transhumanismo latinoamericano crítico tendría que ir más lejos. No bastaría con decir “mejora para todos”; habría que discutir qué significa mejorar, quién lo define y con arreglo a qué concepción de humanidad y de vida buena. El primer criterio normativo que lo distinguiría sería el de relacionalidad. Frente al individualismo del tecnoprogresismo liberal, una versión latinoamericana crítica tendría que sostener que ninguna transformación humana puede evaluarse solo desde la preferencia individual. Toda mejora afecta tramas de dependencia, comunidad, instituciones, vínculos ecológicos y formas compartidas de existencia. El sujeto no sería concebido como propietario

aislado de su cuerpo y su mente, sino como ser relacional, situado y co-constituido. Esta diferencia cambia todo el vocabulario del debate: el derecho a modificarse ya no puede pensarse al margen del mundo social y ecológico en que esa modificación cobra sentido.

El segundo criterio sería el de justicia estructural. El tecnoprogresismo liberal suele centrarse en el acceso y en las oportunidades. En América Latina eso resulta insuficiente, porque las desigualdades no son solo distributivas; son históricas, coloniales, infraestructurales y territoriales. Por tanto, una innovación no debería juzgarse solo por cuántas personas pueden acceder a ella, sino por si reduce o profundiza relaciones estructurales de subordinación. Este criterio desplaza la pregunta desde “¿quién entra al mercado de la mejora?” hacia “¿qué tipo de orden social produce esta tecnología?”. El tercer criterio sería el de pluralismo civilizatorio. El tecnoprogresismo liberal tiende a asumir una idea relativamente unificada de progreso: más capacidades, más autonomía, más opciones. En cambio, una perspectiva latinoamericana tendría que reconocer que existen múltiples imágenes legítimas del futuro humano. No todas las sociedades desean orientar su horizonte hacia la maximización biotecnológica del individuo. Algunas priorizan equilibrio ecológico, cohesión comunitaria, cuidado, reciprocidad o reparación histórica. Por eso, el criterio normativo no sería la expansión abstracta de opciones, sino la compatibilidad de una tecnología con

distintos modos de vida y distintos horizontes de dignidad. Esta pluralidad libera al transhumanismo de su tendencia a presentarse como destino único de la especie y lo obliga a entrar en una conversación más amplia sobre qué futuros queremos y por qué.

El cuarto criterio distintivo sería el de sostenibilidad ontológica y ecológica. El tecnoprogresismo liberal suele pensar la tecnología como ampliación de libertad humana, aunque no siempre interroga suficientemente la relación entre mejora humana y mundo compartido. Una perspectiva latinoamericana, influida por el Buen Vivir y por las críticas a la separación moderna entre humanidad y naturaleza, exigiría que toda promesa de *enhancement* sea evaluada también por su impacto en ecosistemas, territorialidades y relaciones entre humanos y no humanos. El punto ya no sería solo cuánto podemos transformar el cuerpo, sino si esa transformación profundiza formas destructivas de habitar el mundo. El quinto criterio sería el de vulnerabilidad reconocida. El tecnoprogresismo liberal suele concebir la vulnerabilidad como déficit a reducir. Un transhumanismo críticamente reformulado debería reconocer que no toda vulnerabilidad es un fallo técnico. La dependencia, la finitud, la exposición al otro y el cuidado son dimensiones constitutivas de la vida humana. Esto no impide combatir enfermedades o sufrimientos evitables, pero sí impide convertir toda fragilidad en objeto automático de optimización. Normativamente, esto marca una diferencia

profunda: no todo lo técnicamente superable debe ser existencialmente abolido. Finalmente, el sexto criterio sería el de deliberación democrática sustantiva. El tecnoprogresismo liberal tiende a confiar en una mezcla de innovación de mercado, regulación estatal y elección informada del consumidor-ciudadano. El modelo que aquí se propone requeriría algo distinto: que las definiciones sobre mejora, cuerpo, cognición, longevidad o integración humano-máquina sean objeto de deliberación colectiva robusta, no solo de consentimiento individual. No porque la autonomía personal deje de importar, sino porque las tecnologías transhumanistas rediseñan condiciones comunes de existencia y, por tanto, exceden claramente el marco de la elección privada.

En términos sintéticos, el modelo institucional viable sería uno de soberanía tecnocientífica pública, cooperación regional, regulación ética fuerte y deliberación democrática, articulado con sistemas de salud, educación e investigación. Y sus criterios normativos distintivos frente al tecnoprogresismo liberal serían: relacionalidad, justicia estructural, pluralismo de futuros, sostenibilidad ecológica, reconocimiento de la vulnerabilidad y primacía del bien común sobre la mera elección individual. Esto permitiría hablar no de un simple transhumanismo “más inclusivo”, sino de un transhumanismo posliberal, capaz de incorporar las críticas latinoamericanas a la colonialidad, a la subordinación tecnológica y a la reducción del ser humano a un paquete de

funciones mejorables. Bajo esta luz, la contribución del Sur no sería meramente correctiva, como si se limitara a suavizar los excesos del proyecto original. Sería constitutiva: ofrecería otra gramática para pensar la técnica, otra semántica para imaginar la mejora y otra ontología para comprender lo humano. En vez de un futuro organizado por la competencia entre sujetos amplificados, podría pensarse un porvenir orientado por la convivencia entre diferencias, por la reparación de desigualdades y por una técnica puesta al servicio de la cohabitación respetuosa entre cuerpos, comunidades y territorios.

Todo esto permite responder, al menos provisionalmente, a la pregunta que abrió esta sección. Sí, América Latina puede tener futuro dentro del debate transhumanista, aunque no como simple apéndice de un proyecto ajeno ni como periferia que adopta tardíamente los sueños de otros. Su verdadera posibilidad reside en intervenir críticamente en la definición del propio transhumanismo, obligándolo a atravesar la prueba de la justicia, de la relacionalidad, del territorio, de la comunidad y de la pluralidad de futuros. El gran aporte latinoamericano consiste en recordar que mejorar la vida no equivale necesariamente a abandonar la fragilidad, la dependencia o el vínculo con el mundo común. A veces mejorar significa aprender a cuidar mejor, a convivir mejor, a transformar sin devastar, a ampliar capacidades sin erosionar los lazos que sostienen la existencia. En esa dirección puede pensarse, con

cautela, la posibilidad de un transhumanismo ético y democrático desde el Sur. No sería un proyecto de superación solitaria, sino de coexistencia mejorada; no estaría orientado por la supremacía sobre otros cuerpos y otros territorios, sino por la capacidad de habitar con respeto y adaptación mutua; no se mediría por la distancia recorrida respecto de lo humano, sino por la profundidad del nosotros que logra construir.

## **Conclusiones**

El análisis desarrollado en este capítulo permite concluir que el transhumanismo no constituye únicamente una corriente especulativa sobre el futuro, sino un proyecto tecnocultural con capacidad real para intervenir en la definición contemporánea de lo humano. Sus promesas de expansión biológica, cognitiva y vital están hoy sostenidas por redes institucionales, inversión científica y discursos de innovación que buscan materializar una determinada imagen del porvenir. En este contexto, la cuestión decisiva ya no consiste solo en preguntarse si estas tecnologías serán posibles, sino en examinar qué concepción de humanidad organizan, qué desigualdades pueden profundizar y qué formas de vida privilegian como deseables. Desde América Latina, esta discusión adquiere una densidad particular. La región no solo enfrenta condiciones estructurales de dependencia tecnológica, sino que también dispone de una tradición intelectual capaz de interrogar críticamente la

universalidad de las narrativas dominantes sobre el futuro. El pensamiento de la liberación, la crítica de la colonialidad, la reflexión sobre territorio y comunidad, y especialmente el Buen Vivir, ofrecen herramientas para desplazar el debate desde la lógica de la optimización individual hacia una reflexión más amplia sobre justicia, relacionalidad, sostenibilidad y vida compartida. En esa medida, la contribución latinoamericana no consiste simplemente en oponerse al transhumanismo, sino en mostrar que la técnica puede pensarse desde otros fines y otras formas de habitar el mundo, sin reducir la vida humana a rendimiento, control o ampliación indefinida de capacidades.

El capítulo ha mostrado, además, que una reformulación ética y democrática del transhumanismo solo sería posible si este abandona su núcleo elitista, corporativo y tecnocrático. Ello exige transformaciones tanto institucionales como normativas. En el plano institucional, implicaría fortalecer una arquitectura pública de innovación tecnocientífica, cooperación regional, regulación ética robusta y deliberación democrática, de modo que la investigación no quede subordinada exclusivamente al mercado ni a la promesa empresarial de la disrupción. En el plano normativo, supondría desplazar el énfasis desde la mera ampliación de capacidades hacia criterios como relacionalidad, justicia estructural, pluralidad de futuros, sostenibilidad ecológica y reconocimiento de la vulnerabilidad. Bajo esta luz, América Latina puede tener

futuro dentro del debate transhumanista, aunque no como apéndice de un proyecto ajeno, sino como espacio crítico capaz de intervenir en su resignificación. Su aporte más importante consiste en recordar que mejorar la vida no equivale necesariamente a abandonar la fragilidad, la dependencia o el vínculo con el mundo común. A veces, mejorar significa aprender a cuidar mejor, a convivir mejor, a transformar sin devastar y a ampliar capacidades sin erosionar los lazos que sostienen la existencia. Frente a la promesa de trascender los límites humanos mediante innovación tecnológica, quizá el desafío más urgente no consista en ir más allá de lo humano a cualquier costo, sino en aprender a vivir mejor juntos dentro de los límites que hacen posible una vida compartida.

## Referencias

- Acosta, A. (2010). *El Buen Vivir en el camino del postdesarrollo: Una lectura desde la Constitución de Montecristi*. Fundación Friedrich Ebert.
- Acosta, A. (2013). *El Buen Vivir. Sumak Kawsay: Una oportunidad para imaginar otros mundos*. Icaria.
- Allied Market Research. (2022). *Transhumanism market forecast, 2022–2030*.
- Bostrom, N. (2005). A history of transhumanist thought. *Journal of Evolution and Technology*, 14(1), 1–25.
- Bratton, B. (2021). *The revenge of the real: Politics for a post-pandemic world*. Verso.



- Dussel, I. (2004). *Desigualdades sociales y desigualdades escolares en la Argentina de hoy: Algunas reflexiones y propuestas*. FLACSO, Facultad Latinoamericana de Ciencias Sociales, Sede Argentina.
- Echeverría, B. (1998). *La modernidad de lo barroco*. Ediciones Era.
- Escobar, A. (2018). *Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds*. Duke University Press.
- Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. En S. Jasanoff & S.-H. Kim (Eds.), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power* (pp. 1–33). University of Chicago Press.
- Kurzweil, R. (2006). *The singularity is near: When humans transcend biology*. Penguin.
- Kusch, R. (2000). *América profunda*. Biblos.
- Mignolo, W. D. (2005). *La idea de América Latina: La herida colonial y la opción decolonial*. Gedisa Editorial.
- Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. PublicAffairs.
- Quijano, A. (1992). Colonialidad y modernidad/racionalidad. *Perú Indígena*, 13(29), 11–20.
- Santos, B. de S. (2010). *Descolonizar el saber, reinventar el poder*. Trilce.
- UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

# **La eficiencia como categoría de la acción sostenible. Sobre la genealogía filosófica y la ética política de la indulgencia más allá del paradigma de la eficiencia**

*Efficiency as a Category of Sustainable Action. On the Philosophical Genealogy and the Political Ethics of Forbearance beyond the Efficiency Paradigm*

**David Manolo Sailer<sup>13</sup>**

---

<sup>13</sup> Albert-Ludwigs-Universität Freiburg

## Resumen

Este artículo analiza críticamente el concepto de ‘eficiencia’ como categoría central para la acción sostenible. Sostiene que la eficiencia no es una métrica técnico-neutral, sino un dispositivo de racionalidad con una genealogía filosófica ligada al proyecto moderno de dominio sobre la naturaleza. A través del análisis del ‘efecto rebote’ (rebound effect), se demuestra la insuficiencia del paradigma de la eficiencia para lograr reducciones absolutas en el consumo de recursos. El texto propone una revisión de la relación medio-fin, abogando por un desplazamiento del enfoque desde la eficiencia de los medios hacia la resiliencia del fin. Se concluye que una ‘nueva honestidad’ debe reconocer los límites de la mitigación e integrar la adaptación, la resiliencia y la suficiencia como pilares de una ética climática robusta para afrontar las pérdidas inevitables en el Antropoceno.

**Palabras clave:** Eficiencia; efecto rebote; adaptación; resiliencia; ética climática.

## Abstract

This article critically analyzes the concept of ‘efficiency’ as a central category for sustainable action. It argues that efficiency is not a technically neutral metric but a rationality dispositive with a philosophical genealogy linked to the modern project of disciplining nature. Through the analysis

of the ‘rebound effect’, the insufficiency of the efficiency paradigm to achieve absolute reductions in resource consumption is demonstrated. The text proposes a revision of the means-end relationship, advocating for a shift in focus from the efficiency of the means to the resilience of the end. It concludes that a so-called ‘new honesty’ must acknowledge the limits of mitigation and integrate adaptation, resilience, and sufficiency as pillars of a robust climate ethic to face the inevitable losses of the Anthropocene.

**Keywords:** Efficiency; rebound effect; adaptation; resilience; climate ethics.

## Introducción

La práctica discursiva actual de la *digitalized sustainability*<sup>14</sup> entrelaza – abstraída de estrategias de

---

<sup>14</sup> Por *digitalized sustainability* (sostenibilidad digitalizada) se pueden entender actualmente todas aquellas tecnologías digitales – desde la Inteligencia Artificial hasta el Big Data, el Internet de las Cosas (IoT) o las Smart Grids – que se emplean como palancas para alcanzar objetivos de sostenibilidad. Esto abarca tanto la optimización de procesos ecológicos mediante el control digital, el monitoreo y las ganancias de eficiencia, como la reflexión crítica sobre los costes consecuentes materiales, energéticos y sociales de las infraestructuras digitales o de dichas smart solutions; como tal, el discurso se sitúa en la tensión entre los polos de debate de los enfoques de sostenibilidad tecno-optimistas y los orientados al postcrecimiento.

afrontamiento individuales – dos de los discursos de transformación más influyentes y dominantes a nivel social en la actualidad: la transformación sostenible por un lado, y la innovación tecnológica a través de la digitalización por el otro. El diagnóstico muestra que ambos metadiscursos, en su forma y tipo de relación de mediación entrelazada, no siempre están armónicamente vinculados, sino que están estructuralmente marcados por una heurística común: la *racionalidad instrumental de la eficiencia* (como categoría de acción). La eficiencia se enfoca principalmente en la *relación* – en términos simples: la relación entre *input* y *output* – y opera, por definición, dentro de una logicidad técnico-económica que, si bien presupone la intencionalidad como establecimiento de un fin, rara vez la reflexiona. Como *modus operandi*, esta reducción metodológica es en sí misma eficiente y efectiva, y por lo tanto (económicamente) “útil”, pero solo mientras los objetivos no sean controvertidos; sin embargo, se convierte en una trampa tan pronto como la legitimidad normativa de los objetivos – y con ello los grados de consecución de los mismos y los corredores de acción del obrar colectivo – se ponen o son puestos en tela de juicio.

Ahora bien, en una era en la que “der menschliche Naturumgang ob [einer] signifikanten Überbelastung anthropogener Naturkontrolle defizitär zu werden droht”

(Sailer, 2022, p. 17)<sup>15</sup>, es preciso conectar precisamente esas dinámicas de transformación y los megatemas discursivos globales de la sostenibilidad y la digitalización. Dos ideas analíticas parecen centrales de antemano para no errar en las premisas de la pregunta: primero, el *vacío categorial* del concepto de eficiencia *en su incrustación semántica*, y segundo, la *economía política de su implementación*. Como se expondrá más adelante, el concepto de eficiencia en sí mismo es instrumental: (todavía) no dice nada sobre si el objetivo perseguido es éticamente imperativo, socialmente justo o ecológicamente compatible en su teleología vectorial. La eficiencia como mera relación puede, por lo tanto, *muy bien* fomentar estados finales altamente destructivos o normativamente inapropiados, siempre que el objetivo mismo no sea objeto de problematización y reflexión críticas. En segundo lugar, su integración en los mecanismos de mercado y de gobernanza estatal conduce inevitablemente a complejas dependencias de trayectoria sociopolíticas, que se manifiestan, entre otras cosas, en esquemas de subvenciones, programas de infraestructura, prácticas de inversión o estándares regulatorio-tecnológicos, los cuales favorecen aquellas soluciones que a corto y mediano plazo parecen ‘eficientes’ y, por consiguiente, político–económicamente convenientes, pero que a largo plazo también pueden – no

---

<sup>15</sup> “... la relación del hombre con la naturaleza amenaza con volverse deficitaria debido a una sobrecarga significativa del control antropogénico de la naturaleza” (Sailer, 2022, p. 17).

siempre – intensificar las fragilidades. En términos más concretos, ante la escalada de la crisis climática, que ya está causando daños en parte irreversibles (sequías, inundaciones, fenómenos meteorológicos extremos, pérdida de biodiversidad), una ética que se basa únicamente en la mitigación a través de la eficiencia<sup>16</sup> es solo sostenible de forma limitada. O dicho sin rodeos: el mundo debe prepararse para el hecho de que no todas las medidas de mitigación tendrán éxito y que los impactos negativos del cambio climático se sentirán dolorosamente; por lo tanto, la adaptación – en el sentido de una aceptación activa de la acción y una configuración proactiva de las vulnerabilidades – gana en urgencia ética y de *realpolitik*. Sin embargo, esto requiere una “nueva honestidad” y una previsión indulgente similar, es decir, que *no todo saldrá bien* y que nuestros modelos sociales *tendrán* que ajustarse a las consecuencias inevitables del cambio climático y sus impactos negativos – no deberían, ya que el hecho de la resiliencia real de los espacios habitados se convierte en una práctica política realista. Y esta “nueva honestidad” exige un desplazamiento del horizonte normativo y del corredor de objetivos, desde

---

<sup>16</sup> Al mismo tiempo, a fecha de 2025, debe constatarse que la situación geopolítica actual demuestra de manera dramática que, mientras algunos actores apuestan por tecnologías verdes y una economía sostenible, se observa simultáneamente una reacción adversa (*backlash*) que comienza a socavar los esfuerzos por una transición energética global.



una fijación unilateral en la eficiencia hacia una concepción plural y orientada a la resiliencia de la sostenibilidad.

El objetivo principal de este artículo es demostrar que la eficiencia, lejos de ser un mero instrumento técnico-neutral, opera como un dispositivo normativo que perpetúa la racionalidad extractivista moderna. A través del análisis del “efecto rebote” (*rebound effect*), se evidenciará empíricamente la insuficiencia del paradigma optimizador para lograr una reducción absoluta en el consumo de recursos. Ante el fracaso de este “canto de sirena de la optimización”, la tesis normativa que defiende este texto es la urgencia de abandonar la primacía de la eficiencia de los medios. En su lugar, se propone un desplazamiento hacia la resiliencia del fin, la suficiencia sistémica y lo que denominaremos una “actitud de indulgencia”. Esta nueva configuración ética y política resulta indispensable para fundar una “nueva honestidad” capaz de afrontar las pérdidas irreversibles en la era del Antropoceno.

En última instancia, la eficiencia no es una categoría técnica neutral en cuanto a valores, sino un dispositivo de racionalidad que ha crecido cultural e ideohistóricamente, que privilegia ciertas relaciones con el mundo y luego las ‘pone en práctica’. La situación geopolítica actual – caracterizada por conflictos en escalada, una creciente precariedad social y una crisis climática que se agudiza – plantea a la humanidad inmensos desafíos, cuya acumulación

cuestiona radicalmente las narrativas de progreso familiares de la modernidad europea. En el centro de estas narrativas se encuentra desde hace tiempo una promesa: la promesa de la eficiencia, entendida como la capacidad técnica y económica de realizar cada vez más con cada vez menos en intervalos de tiempo cada vez más cortos. Sin embargo, en la era del Antropoceno incipiente, esta promesa se revela como un principio rector ambivalente, a menudo engañoso y de dos caras. Las crecientes curvas de temperatura, la ausencia de puntos de inflexión estabilizadores más allá de las esquivas metas de 1,5 grados y la emergencia de múltiples efectos de retroalimentación sugieren que la ‘eficiencia’ por sí sola no es ni sistémicamente suficiente ni normativamente deseable en todos los casos<sup>1</sup>.

La expresión quizás más evidente de esta aporía contemporánea parece ser el llamado efecto rebote (Berkhout et al., 2000; Brockway et al., 2021; Frondel et al., 2012; Galvin, 2015; Gillingham et al., 2016), según el cual las ganancias en eficiencia son ‘compensadas’ por ajustes en el comportamiento, un uso más intensivo o también una escalada estructural de la demanda; es decir: los ahorros derivados del progreso tecnológico no conducen necesariamente a una reducción absoluta del consumo de recursos o energía, sino que, paradójicamente, a veces generan nuevos impulsos de demanda de tal naturaleza que pueden neutralizar el efecto neto esperado o incluso invertirlo. Sin embargo, esta observación empírica (Santarius,

2016; Santarius et al., 2016; Sorrell, 2007; Sorrell y Dimitropoulos, 2008; Vivanco et al., 2022) es posiblemente solo la punta de un ‘iceberg discursivo’, ya que la progresiva digitalización y la automatización (digitalmente inteligente) conectada a ella de procesos intelectuales, así como económico-logísticos, conducen a una especie de segunda forma de rebote menos observada: una inflación digital-epistémica. Más output en forma de datos, cálculos e información no sería entonces, per se, sinónimo de más conocimiento en el sentido de claridad epistémica. Y a la luz de estas consideraciones iniciales, surge la pregunta central de este artículo: si la eficiencia – entendida como la heurística primaria de asignación de costos y beneficios, que se mantiene orientada a cálculos de costos y beneficios marginales – es realmente una categoría adecuada para la acción sostenible. O si se trata más bien de una forma de pensamiento que oculta las causas sistémicas de una crisis, perpetúa las dependencias institucionales de trayectoria que se mueven en un curso de deficiencias progresivas y, por lo tanto, obstaculiza los esfuerzos de transformación fundamentales.

## **La engañosa promesa de la optimización en el Antropoceno**

Uno de los mitos más persistentes de la modernidad es que el ser humano, en virtud de su *conditio humana*, se realiza a sí mismo *en el progreso* y se revela a sí mismo. El

*progreso*, esta salida de un comienzo autoimpuesto (cf. Sailer-Wlasits, 2020) – entendido como un movimiento contingente y progresivo en dirección a una autoactualización cada vez mejor y a una perfección colectiva –, alberga sin embargo un alto potencial de deficiencia y, a pesar de todo, fue rearticulado a lo largo de los siglos no solo como una forma tecnológica, sino como un telos civilizatorio. Aquí, el “progreso” es un motivo constitutivo de la modernidad, en la medida en que esta autocomprensión teleológica – y por lo tanto, la autoafirmación inmanente en el mundo – ha generado la racionalización, tecnificación y explotación económica de las relaciones entre el hombre y la naturaleza. *Onerosi sumus mundo* – “nos hemos convertido en una carga para la tierra”<sup>17</sup> –, ya sabía informar el escritor eclesiástico romano-cartaginés Tertuliano en su obra *De anima* en el siglo II d.C. Y es una frase que no solo logra transmitir el peso normativo específico de todo lo que el ser humano lleva a cabo en el Antropoceno, sino que también sabe señalar con precisión el tríptico (trans)formador del mundo de la relación antropogénica con la naturaleza: una normatividad de la naturaleza, un concepto de naturaleza moderno que se refleja

---

<sup>17</sup> Cf. Tertuliano, *De anima*, XXX, 4: “Es evidente, al menos, que el propio orbe terrestre está cada día más cultivado y es más civilizado que antes. Todo es ya accesible, todo está explorado, todo abierto al comercio; [...] en todas partes hay viviendas, en todas partes poblaciones, en todas partes Estados, en todas partes vida. Somos una carga para la Tierra, apenas nos bastan los elementos, las necesidades se agudizan y por doquier hay quejas, pues la naturaleza ya no quiere sustentarnos.”

en la *sujeción*, así como regímenes de control técnico–amplificadores para la disciplinarización de la naturaleza.<sup>18</sup>

En medio de la acumulación de riesgos en cascada<sup>19</sup> que se amplifican sistémicamente – puntos de inflexión climáticos críticos (*tipping points*), extinción de especies, fragmentación geopolítica, agotamiento de recursos a lo largo de los límites planetarios (Steffen et al., 2009), etc. –, la categoría rectora de la eficiencia se revela, por lo tanto, como provocadoramente bicéfala. En este sentido, parece necesaria una interjección crítico-conceptual para señalar la amalgama semántica de “eficiencia” y “optimización” en numerosos debates contemporáneos sobre sostenibilidad, pues a menudo parece que “eficiencia” es semánticamente ambivalente: como puro rendimiento, describe una relación neutral; histórico–políticamente, sin embargo, se la suele cargar como una *categoría cuasi–normativa* que vincula las

---

<sup>18</sup> Sobre este tríptico, cf. Sailer (2023); específicamente sobre el concepto de *disciplinarización de la naturaleza*, originalmente Sailer (2022).

<sup>19</sup> Cf. IPCC (2022, pp. 82 y ss.): Los riesgos en cascada comprenden las manifestaciones de todos aquellos fenómenos del cambio climático que el IPCC considera consecuencias irreversibles a lo largo de los límites planetarios o efectos dominó debidos a puntos de inflexión climáticos críticos (*tipping points*). Además de los impactos compuestos (*compound*) y en cascada (*cascading*), también deben considerarse los impactos transfronterizos (*transboundary impacts*) del cambio climático en los seres humanos y los ecosistemas, ya que estos surgen de las complejas interacciones entre diferentes peligros climáticos (*climate hazards*).

tendencias de mejora técnica con la corrección moral. Descrito con mayor precisión, el ‘canto de sirena de la eficiencia’ se nos presenta superficialmente como un esfuerzo orientado a la optimización, la mejora inducida por la tecnología y, por tanto, a la configuración positiva del mundo (y del entorno) vital. Pero antes de emprender la reconstrucción genealógica de la lógica de la eficiencia en su génesis moderno-tecnofilosófica, ya la semántica histórica muestra – “eficiencia” del latín *efficere* (*ex* – hacia afuera, *facere* – hacer, producir, causar) – que eficiencia es en el fondo un concepto originalmente *relacional*: inicialmente designaba la *eficacia* de una acción o una causa, es decir, la capacidad específica de producir algo realmente, y menos la minimización económica del esfuerzo. Fue solo en la filosofía de la temprana modernidad, en el contexto discursivo de la causalidad empírico–mecanicista, que el acento del concepto se desplazó hacia la *causa efficiens*, es decir, la *causa eficiente* – a diferencia de la *causa finalis* aristotélica – , con lo cual se dio paso a la eficiencia en el sentido moderno, que desde entonces se refirió más fuertemente a la producción técnica, con ello a la relación medio-fin y, por consiguiente, más tarde, en el curso de la Revolución Industrial, a la relación entre los factores de *input* y *output*. Ahora, pues, mide los grados de rendimiento, la relación entre esfuerzo y resultado, y su heurística tecnicista aparece casi como garantía y medida de la calidad – y, por tanto, de la explotabilidad económica. En cuestiones de naturaleza, esto es una gradación de la

utilización y el consumo, y dice, por tanto, *cuán bien* se puede alcanzar un determinado “objetivo”, un *output* deseado con un uso de recursos (idealmente) optimizado en términos de utilidad marginal y costo marginal.

Aunque el concepto de *optimizar* parece traslucirse aquí, a la idea le falta el núcleo normativo, concretamente en el sentido etimológico original de *optimus*, es decir, la idea de una mejora cualitativa. En otras palabras: ‘eficiencia’ como categoría de acción para una relación manejable con el mundo no dice todavía si un objetivo es en sí mismo legítimo o siquiera deseable.<sup>20</sup> Aplicado a las metanarrativas histórico-culturales – y el discurso de la sostenibilidad no es otra cosa –, esto significa que tan pronto como ‘eficiencia’ se equipara discursivamente con ‘progreso’, un concepto originalmente descriptivo-instrumental experimenta una *recodificación casi eufemística*. Se convierte en un *token* de una narrativa con connotaciones positivas que margina discursivamente objetivos alternativos como la suficiencia, la justicia o la integridad planetaria. Y positivamente connotado, en la medida en que, orientado por el significado original de la palabra, la *optimización* remite a una categoría de valor de *mejora* teleológicamente dirigida o incluso – pensado idealistamente – de *perfeccionamiento*, y *precisamente en*

---

<sup>20</sup> No es posible realizar aquí una discusión sistemática de las consecuencias deontológicas de una heurística de objetivos con anclaje teleológico, la cual queda reservada para reflexiones posteriores.

eso no es una relación ‘neutral’. Sin embargo, esta fusión semántica es políticamente efectiva, es decir, es afirmativa en su estructura narrativa fundamental y en su retórica cotidiana, porque es favorable al progreso. No obstante, funciona también como una fórmula de confirmación normativa para tecnologías y prácticas sociales que a veces reproducen desigualdades sociopolíticas, cargas ecológicas y dependencias geopolíticas. Al mismo tiempo – y esto también lo muestran las *trayectorias* discursivas de la modernidad entre el progreso y la decadencia (cf. Kerkmann y Sailer, en prensa) –, el concepto de optimización a veces oculta y encubre la contingencia y selectividad de los objetivos establecidos cultural y colectivamente, al pretender aspirar siempre a “lo mejor” y “lo más alto” en aras del bien común. La norma de la eficiencia oculta, además, una cierta ceguera a la escala: lo que a nivel de procesos, tecnologías o procedimientos individuales en un sistema dispositivo (cf. Foucault, 2004) parece *eficiente* y se desarrolla de manera *funcional*, puede, a nivel agregado – debido, entre otras cosas, a disfuncionalidades colectivas de la teoría de juegos –, conducir paradójicamente a mayores cargas y, por tanto, *summa summarum*, a una carga adicional para el sistema. La metanarrativa de la eficiencia pasa por alto la emergencia y la no linealidad de los sistemas socio-ecológicos, y con ello también aquellas interacciones que solo se hacen visibles a mayor escala y a lo largo de períodos de tiempo más largos, posiblemente incluso intergeneracionales.



De lo anterior se desprende que, siempre que se produce una amalgama conceptual-discursiva de eficiencia y optimización, se lleva a cabo una *carga normativa* de un concepto de *efecto* originalmente *neutral en cuanto a valores*. En este contexto, es necesaria una reconstrucción genealógica de la lógica de la eficiencia, pues solo ella muestra que la eficiencia no es meramente un cálculo técnico o una simple medida de racionalización, sino que, en la medida de su aplicabilidad, también se codifica como normativamente positiva, es decir, como una alternativa deseable o incluso normativamente requerida (por ejemplo, frente a ideas de redefinición de la “relación hombre-naturaleza”). Como se puede ver, son aporías conceptuales en las curvas límite de la categoría rectora de la sostenibilidad, de modo que en la visión de conjunto de las determinaciones surge una idea trascendental: si la eficiencia representa *en absoluto* una categoría adecuada para la acción sostenible. Pues en una era en la que la relación del hombre con la naturaleza amenaza con volverse deficitaria debido a una sobrecarga significativa del control antropogénico de la naturaleza, parece necesario considerar la *vitalidad natural* y (“la permanencia de la vida humana auténtica en la Tierra”, Hans Jonas<sup>21</sup>) de una manera

---

<sup>21</sup> Al recurrir al imperativo ecológico formulado por Hans Jonas, no apelamos a una noción abstracta, sino a una premisa de supervivencia muy concreta: la innovación técnica y la acción política deben garantizar la permanencia material de una vida humana auténtica en la Tierra. En el debate sobre la sostenibilidad, esto significa que el simple “poder hacer” tecnológico (la eficiencia) debe subordinarse éticamente a la preservación

más integral: Los discursos dominantes sobre sostenibilidad revelan inicialmente el afán de superación inherente a todas las dinámicas de transformación abiertas a la tecnología, al ofrecer prometedoras promesas de reparación. Sin embargo, la mayoría de las tecnologías de reparación, como intervenciones ambientales de los llamados *technological fixes*, se asemejan más a medidas para corregir daños ya ocurridos en los contextos naturales (cf. Sailer, 2023, p. 52).

### **Sobre la genealogía de la eficiencia. La disciplinarización de la naturaleza como proyecto filosófico de optimización**

Los problemas lógico-conceptuales no eximen de la necesidad de su aplicación práctica; es necesario un desenredo estructural. En otras palabras, la eficiencia ya no debería servir como fórmula de sustitución y término de reemplazo para la optimización, ya que la relevancia ético-ecológica de la acción (u omisión) colectiva no puede derivarse únicamente de las relaciones de rendimiento. A la cuestión de la mejora eficiente de los medios debe añadirse inevitablemente la de las propias relaciones medio-fin, lo que constituye la “puerta de entrada” filosófica para una investigación genealógica de la *lógica de la eficiencia*. Sin embargo, en este contexto hay que tener en cuenta que *esta*

---

del futuro. Una tecnología altamente eficiente deja de ser legítima si, por efecto rebote, agota irreparablemente las condiciones de posibilidad de las generaciones venideras, cf. Jonas (1984).

no es un subproducto arbitrario de la modernización industrial – respectivamente, de la industrialización capitalista-occidental –, sino el resultado de un proyecto cultural-histórico-antropológico más profundo que, con el comienzo de la modernidad, alcanzó su apogeo intelectual en una semántica de la naturaleza orientada al dominio y que solo se volvió *técnicamente aplicable* con la Revolución Industrial.

### *Génesis filosófica moderna. La naturaleza como objeto de dominio*

No es necesario aquí emprender una nueva exposición del concepto de naturaleza en sí mismo y en su normatividad; sin embargo, sí es preciso señalar que el paradigma científico (natural) racionalista de la temprana modernidad – paradigmáticamente en René Descartes, pero también en la concepción matemático-cosmológica de la naturaleza de Galileo Galilei o en la lógica del derecho natural de Thomas Hobbes – realiza un cambio trascendental, una ruptura en el pensamiento según la μεταβολή (*metabolē*), el cambio súbito y repentino (cf. Gadamer, 1999, p. 42): la naturaleza ya no se experimenta como un *mundo circundante fenoménico* o un *entorno vital animado*, sino como una *res extensa*, una materia cuantificable, modelable y controlable o disciplinable. Aún más trascendental es que con esto se establece una clara separación: el sujeto cognoscente se enfrenta a la naturaleza objetivada, cuyo valor se deriva ahora

de su utilidad *objetiva* y, por tanto, de su *explotación práctica*. La *metabolē* resuena aquí como el *control* que madura como medio de elección, ya sea según la metodología empírica o los métodos de procedimiento económico-técnicos: la metáfora de que el hombre es “dueño y poseedor” de la naturaleza (*maîtres et possesseurs de la nature*; Descartes, 2011, VI) – una idea que, por supuesto, sigue vinculada al mandato de la creación a la luz del *dominii terrae* – funciona desde entonces como la narrativa rectora retórica del dominio moderno de la naturaleza; y la transición semántica del dominio de la naturaleza a su subyugación siempre fue fluida. En esta inclinación ya reside, como es fácil de reconocer, el germen lógico-conceptual, normativo y socio-filosóficamente muy relevante, de la heurística de la eficiencia: pues si la ‘naturaleza’ puede ser pensada como un objeto controlable – uno que se puede cuantificar, medir y dirigir, y por tanto “optimizar” –, su *magia viva* se desvanece. Y esto no significa otra cosa que la secularización moderna de la naturaleza o el dominio de la naturaleza es una relación *en y como jerarquización*. Es una relación de supra y subordinación – concretamente, la del espíritu y la naturaleza – y esta dualidad cartesiana sujeto-objeto constituye, según la *razón instrumental*<sup>22</sup>, el fundamento metafísico y metaético

---

<sup>22</sup> En el presente texto, por razón instrumental se entiende una forma de acción racional que se concentra únicamente en la selección funcional y el uso eficiente de los medios, sin reflexionar sobre la calidad moral o la legitimidad de los fines perseguidos. En la *Teoría Crítica* (Adorno/Horkheimer) se la critica como una racionalidad reduccionista

de esta misma heurística de la eficiencia. Pero si la naturaleza aparece como una aglomeración de recursos útiles, entonces su *buen uso* debe medirse por la maximización de la relación *output-input*, y cualquier pregunta sobre si el objetivo en sí mismo es éticamente imperativo queda fuera del foco de consideración.

El triunfo civilizatorio de antaño de una *disciplinarización* eficiente de la naturaleza – es decir, el aumento del *output* con la simultánea minimización de los gastos de *input*<sup>23</sup> específicos – se vuelve así contra sí mismo y se convierte en una aporía, tan pronto como ya no se aplica a la lógica económica interna, sino a las condiciones planetarias de la vida misma. Dicho de otro modo, surge una primera *paradoja ético-adaptativa* siempre que las sociedades consideran la naturaleza solo en su función de utilidad para la aplicación humana y no como el fundamento vivo de la posibilidad misma de su existencia. La adaptación, es decir, el ajuste a condiciones ecológicas cambiadas, se concibe

---

que considera el mundo exclusivamente desde el punto de vista de su explotabilidad técnica, una actitud que conduce a la instrumentalización de la naturaleza y del ser humano.

<sup>23</sup> Piénsese en este contexto, por ejemplo, en el aumento del rendimiento agrícola por hectárea mediante el uso de semillas de alto rendimiento y la agricultura de precisión controlada digitalmente con un consumo de agua reducido, o también en los procesos de fabricación totalmente automatizados en la producción industrial, así como en las tecnologías de propulsión más eficientes en el transporte, que consumen menos combustible por kilómetro o milla aérea con mayor autonomía.

entonces a menudo como un déficit y no como una racionalización de orden superior a la luz de la transformación; como consecuencia, el pensamiento permanece atrapado en lógicas de reparación, es decir, de *soluciones tecnológicas*, y no en la resiliencia adaptativa. Pero la verdadera tragedia reside ahora en la persistencia de este modelo de pensamiento, que modela proyectos de futuro vectoriales y maduró desde la pretensión de dominio hasta la lógica de optimización cuantificada-industrializada. ¿O cómo sería posible de otro modo que la eficiencia se nos presente hoy como la forma economizada de la pretensión de dominio? Sin embargo, la idea de que mediante refinamientos tecnológicos siempre nuevos e inteligencias de control digitales el mundo podría hacerse ‘sostenible’ sin tocar las coordenadas fundamentales socio-estructurales – y es decir, antropológico-culturales – de ese tríptico semántico de “*aprovechamiento – utilización intensiva – sobreexplotación*”<sup>24</sup>, sigue siendo un noble deseo. Por lo tanto, improbable. Es una forma de pensar que lleva en sí rasgos de un científicismo tardío; y es una *fe silenciosa*, en forma de un consenso fundamental tácito de normatividad,

---

<sup>24</sup> Mientras que por *aprovechamiento* (*Nutzbarmachung*) se entiende principalmente el proceso de poner los recursos a disposición humana, los términos *consumo/utilización intensiva* (*Vernutzung*) deben indicar una integración duradera (a veces también de forma permanentemente agotadora) en los ciclos de producción, y el de *explotación/sobreexplotación* (*Ausnutzung*), un agotamiento máximo asumiendo daños sistémicos irreparables.

en una especie de omnipotencia de la razón técnica y del progreso vectorial-técnico, al que se le transfiere una tarea de solución grande y de pesada responsabilidad, incluso cuando la dimensión ecológica y existencial de la crisis parece escapársele cada vez más.

### *La eficiencia como patología sociocultural en la era de la crisis climática*

Todas estas líneas de desarrollo y discursos no son arbitrarios, sino el resultado de todo lo que *se ve*, es decir, lo que comienza a perfilarse en y sobre el Antropoceno. Es, en cierto modo, una consecuencia del *vacío teleológico* en el sistema social y en la lógica de la eficiencia: pues dondequiera que el objetivo mismo no se reflexiona y se articula, el criterio de evaluación discursivo-normativo se desplaza, con una necesidad casi automática, hacia la perfección de los medios. Y así, también los movimientos de sostenibilidad contemporáneos, que se presentan con gusto en un tono apelativo como contramodelos verdes a las dinámicas de consumo modernas, están paradójicamente colonizados de principio a fin por la lógica de la eficiencia, ya que todavía no abordan el sustrato de la constante antropológica fundamental intramundana que debe ser tratada: el *acceso al mundo bajo el primado de la función*.

Vemos: solo cuando se logre pensar desde esta base y constante fundamental, se abrirá – posiblemente – un nuevo

corredor discursivo que no necesita operar con cargas y sobrecargas normativas, evitando así el peligro de que la absolutización de una posición (la propia) se convierta en dogmatismo y, por tanto – como tantas veces en la historia cultural de la humanidad – en la estigmatización de todos los que piensan diferente. Pues el *acceso al mundo bajo el primado de la función* es, en ocasiones, la correspondencia especular del concepto de eficiencia como heurística de sistema de una civilización agotada. Es ante una crisis climática que se agudiza que esta patología se vuelve particularmente visible, porque es dolorosamente perceptible: las innovaciones tecnológicas, que se promocionan como soluciones – ya sean ahorros inteligentes o ganancias de eficiencia escalables – son a menudo, no siempre, autodestrucciones diferidas. La lógica de eficiencia subyacente sigue siendo, en su estructura fundamental, una heurística de costo-beneficio y, por lo tanto, no se orienta por cuestiones de imperativo normativo, justicia intergeneracional o cohesión sociocultural sin fronteras, sino por la optimización escalable de rendimientos, desempeños y beneficios. Pero en *contextos vivos* de este tipo, las leyes de la termodinámica solo se aplican de forma limitada; allí, eficiencia significa minimizar la pérdida de energía.<sup>25</sup> En los

---

<sup>25</sup> Es decir: en los sistemas vivos y abiertos, la eficiencia no solo significa la minimización de pérdidas, sino también la capacidad de gestionar los flujos de energía de tal manera que se aseguren la estabilidad y la capacidad de regeneración.



ecosistemas socio-ecológicos, sin embargo, la pregunta por la calidad y los objetivos de esta “reunión” debe examinarse según una *hermenéutica de la vida* (cf. Angehrn, 2023, pp. 237 y ss.). En contraposición, el aplazamiento de las estrategias de adaptación necesarias y requeridas reprime estas cuestiones normativas fundamentales y, a veces, las hace caer en el olvido. Es precisamente en este trabajo de represión estructural donde se manifiesta la patología colectiva del paradigma de la eficiencia. La heurística de la eficiencia – siempre que se utilice de manera estructuralmente inteligente y específica para la aplicación – sigue siendo indispensable en su función pragmática – pues, por supuesto, se necesitan urgentemente nuevas tecnologías –, pero al mismo tiempo es ciega a aquellas cuestiones categoriales fundamentales sin cuya aclaración ninguna civilización puede ser viable a largo plazo, tampoco en el siglo XXI. Por lo tanto, es necesario encontrar un nuevo objetivo para la acción sostenible, uno que, a la luz de una *nueva y radical honestidad*, reconozca la adaptación como un pilar de igual importancia junto a la mitigación y que, por ello, exija una racionalidad de orden superior en la relación con la naturaleza.

Si se siguen estos pensamientos de la teoría a la práctica, la eficiencia ya no se entendería como un objetivo, sino, a lo sumo, como una estrategia condicionada dentro de un sistema social orientado a la resiliencia. De este modo, a través de la comprensión genealógica, se lograría un

desplazamiento operativo en la práctica política, que podría entonces hacerse fructífero a lo largo de tres dimensiones de acción: (1) Cambio de métricas: los criterios de evaluación deben ir más allá de los meros grados de rendimiento, lo que significa que nuevos indicadores de resiliencia (por ejemplo: biodiversidad, capacidad de restauración, velocidad de adaptación) deben medirse y cuantificarse al igual que la eficiencia.<sup>26</sup> (2) Reorientación de la gobernanza: los procesos de toma de decisiones políticas establecerían la *capacidad adaptativa* como un criterio de igual rango junto a otros indicadores económicos y sociales esenciales. Se trataría, en palabras de Niklas Luhmann, de la *capacidad de dirección del sistema social*<sup>27</sup>, y si se quiere tomar literalmente el “proceso de difusión” de la teoría de sistemas *entre sistema y entorno* para la relación hombre-naturaleza, los programas de

---

<sup>26</sup> Esto podría operacionalizarse en diferentes niveles: en los informes de sostenibilidad de las empresas, en los planes de desarrollo urbano con objetivos de infraestructura orientados a la resiliencia o – pensado en términos de política global – en una actualización de los Objetivos de Desarrollo Sostenible (SDGs) de la ONU.

<sup>27</sup> Cf. Luhmann (1998). En términos de la sociología de Niklas Luhmann, nos referimos a la “capacidad de dirección” (*Steuerungsfähigkeit*) del sistema social. Para comprender esto sin caer en jerga sistémica, basta observar que la sociedad moderna está fragmentada y tiende a operar con “puntos ciegos”: la economía o la tecnología optimizan su propio rendimiento interno (su eficiencia), pero a menudo fracasan al percibir el daño ecológico acumulado que provocan en su entorno natural. Aumentar la resiliencia significa, por tanto, forzar a nuestras instituciones a observar, procesar y adaptarse a estas vulnerabilidades externas, en lugar de ignorarlas ciegamente.

fomento o las subvenciones vinculados a la robustez a largo plazo podrían ser decisivos. (3) Re-narración sociocultural: en conexión con esta dimensión socio-estructural, a nivel de la evolución de las ideas (que constituye la contraparte, la semántica, de la diferenciación estructural del sistema social), las autodescripciones sociales podrían no necesariamente descartar el mito de un “mañana mejor gracias a mayores grados de eficiencia”, pero al menos complementarlo con una narrativa en la que la adaptación y la limitación de daños, en vista de la justicia intergeneracional y de las *pérdidas y daños* (*Loss and Damage*; cf. Wallimann-Helmer et al., 2019), se consideren una expresión de prudencia racional y de visión de futuro abierta:

Climate Resilient Development integrates adaptation measures and their enabling conditions [...] with mitigation to advance sustainable development for all. Climate resilient development involves questions of equity and system transitions in land, ocean and ecosystems; urban and infrastructure; energy; industry; and society and includes adaptations for human, ecosystem and planetary health. [...] Pathways for advancing climate resilient development are development trajectories that successfully integrate mitigation and adaptation actions to advance sustainable development. (IPCC, 2022, p. 32)

## **El canto de sirena de la optimización o sobre la paradoja del efecto rebote**

Mientras que el discurso sobre la “eficiencia sostenible” a menudo está dominado por la promesa de innovaciones tecnológicas y “soluciones verdes”, la realidad de las crecientes curvas de temperatura y de los esquivos objetivos de 1,5 grados (cf. IPCC, 2023) muestra que estas esperanzas no siempre se cumplen. El llamado *efecto rebote* – el fenómeno de que las ganancias de eficiencia son compensadas (Berkhout et al., 2000; Brockway et al., 2021; Galvin, 2015; Gillingham et al., 2016; Santarius, 2016; Sorrell, 2007; Vivanco et al., 2022) por un aumento del consumo y del uso de recursos, anulándose así en el saldo total – ilustra de manera ejemplar que la sostenibilidad *no puede alcanzarse únicamente mediante la optimización*. Las tecnologías digitales y la transformación digital como tendencia macrosocial histórica se presentan como soluciones a los problemas de sostenibilidad generados por la sociedad. Siguiendo una comprensión más detallada del concepto de *technological fix*, las tendencias de innovación y digitalización abiertas a la tecnología muestran, ante todo, el afán de superación inherente a estas estrategias: desde enfoques de metabolismos urbanos, pasando por la economía circular verde y la producción sostenible de hidrógeno, hasta la controvertida cuestión de las emisiones negativas – como las de la Eliminación de Dióxido de Carbono o la Eliminación de Gases de Efecto Invernadero (Greenhouse Gas Removal

GGR) –, las soluciones tecnológicas innovadoras ofrecen prometedoras promesas de reparación. Sin embargo, la mayoría de las “tecnologías de reparación”, como intervenciones en la naturaleza, se asemejan más a medidas para corregir “daños ya ocurridos en los contextos naturales” (Böhme, 1988, p. 15). Si se aplican estas observaciones a la *promesa de la eficiencia*, pronto se revela la paradoja estructural inherente a la misma: la configuración efectiva de un subsistema puede desestabilizar el sistema en su conjunto, tan pronto como esta se inserta en complejos bucles de retroalimentación. Este fenómeno, que puede describirse como el efecto rebote – o la *paradoja de Jevons*<sup>28</sup> –, apunta a una problemática más profunda que se basa, aproximadamente, en que los enfoques de reparación instrumental desconocen las dinámicas emergentes de los sistemas no lineales y/u orgánicos, al optimizar magnitudes aisladas mientras se ignoran los efectos secundarios sistémicos. Si bien la reducción de la complejidad sirve para la obtención (o generación) metódica de confianza (cf. Luhmann, 2014), bajo ciertas circunstancias también se

---

<sup>28</sup> La llamada *paradoja de Jevons* (nombrada así por el economista británico William Stanley Jevons) describe la forma más extrema del efecto rebote, cuyo impacto total supera el 100 por ciento. Ocurre, por tanto, cuando una mejora tecnológica que aumenta la eficiencia en el uso de un recurso conduce finalmente a un incremento del consumo total de dicho recurso (en lugar de la reducción prevista).

pueden pasar por alto importantes interconexiones en sus efectos o daños consecuentes.

Históricamente, el efecto rebote se puede observar bien en la propia historia de la energía: ejemplos brillantes son la máquina de vapor mejorada de James Watt, con la que pudo reducir drásticamente su consumo de carbón; sin embargo, en lugar de generar ahorros, esto condujo a una explosión del consumo de carbón, porque la tecnología ahora podía utilizarse masivamente (cf. Jevons, 1865). Es un patrón estructuralmente equivalente que se repite en el caso de la bombilla y las lámparas LED; estas últimas reducen considerablemente el consumo de energía por unidad de luz, pero la disminución de los costos y las nuevas aplicaciones de iluminación de alta potencia y continua impulsaron el consumo total neto al alza. Algo similar se observa también en los electrodomésticos cotidianos, en la agricultura, en el transporte y la movilidad, y similares (cf. Sorrell, 2007; Herring y Roy, 2007; Schipper y Grubb, 2000) – en resumen: *la eficiencia legitima la expansión*. Precisamente en esto se muestra que las ganancias de eficiencia no solo desencadenan efectos internos sectoriales, sino que impulsan dinámicas de sistema que van mucho más allá del ámbito de aplicación individual. Pero donde el ahorro de recursos abre el espacio para nuevos modos de uso y expansión, también se desplaza la cuestión subsiguiente de la *definición de los daños*; esta debe, por consiguiente, ser pensada de manera igualmente global y sin fronteras, como los bienes colectivos de la

atmósfera y el agua potable. Las estrategias de sostenibilidad que apuestan por las llamadas tecnologías WEM y WAM<sup>29</sup> deben, por lo tanto, adherirse más en el futuro al pensamiento de Jonas, según el cual toda “política es responsable de la posibilidad de la política futura” (Jonas, 1984, p. 209). Las innovaciones digitales y tecnológicas solo pueden ser viables si respetan los límites de la capacidad de carga planetaria, aunque las modelizaciones sigan inevitablemente plagadas de incertidumbres y una capacidad de pronóstico limitada. Además, la diferenciación regional es crucial, ya que los límites de carga de efecto global despliegan su urgencia en contextos locales específicos.<sup>30</sup> Mientras que la era geológica

---

<sup>29</sup> Según la CMNUCC, los escenarios WEM (*with existing measures*) abarcan aquellos que parecen alcanzables con las medidas existentes, mientras que los escenarios WAM (*with additional measures*) comprenden constelaciones que solo pueden lograrse con medidas adicionales, es decir, con tecnologías que aún deben desarrollarse; en este contexto, cabría mencionar la Captura y Almacenamiento de Carbono (CAC) o la Captura y Utilización de Carbono (CUC).

<sup>30</sup> Según Steffen et al. (2009), se identificaron un total de nueve límites planetarios (*planetary boundaries*), de los cuales siete pudieron ser cuantificados: el cambio climático (concentración de CO<sub>2</sub> en la atmósfera en ppm), la acidificación de los océanos, el agotamiento del ozono estratosférico, así como los ciclos biogeoquímicos del nitrógeno y del fósforo; además, el uso global de agua dulce (consumo de aguas superficiales y subterráneas, cf. también agua verde y azul), el cambio en los sistemas de uso del suelo y, finalmente, la velocidad a la que se pierde la diversidad biológica, es decir, la diversidad genética y funcional de una biosfera intacta. La carga de aerosoles atmosféricos no fue cuantificada en el artículo original, aunque estudios más recientes proponen nuevos métodos de medición. En cuanto al uso de agua dulce, actualmente parece

del Holoceno se caracterizó por una relativa estabilidad *de su estado ambiental* (cf. Rockström et al., 2009) – lo que A. Gehlen (1976) reflejó, entre otras cosas, en su idea de la *uniformidad del curso de la naturaleza* –, el concepto y la idea de los *límites planetarios* (*planetary boundaries*) en el Antropoceno marcan la nueva realidad de la fragilidad ecológica. Desde esta perspectiva, se puede formular un *imperativo ecológico* (Hans Jonas), según el cual esta *uniformidad de la naturaleza también debe ser preservada en el futuro* y, por consiguiente, la dinámica de intervención antropogénica debe mantenerse por debajo de los límites de carga planetarios; es decir, sin que la eficiencia se convierta en una justificación para el consumo y el agotamiento irreparable de recursos.

Actualmente, esta lógica de eficiencia y consumo se está desplazando sucesivamente hacia la esfera digital, y la disponibilidad del trabajo intelectual digital y su aplicabilidad (artificialmente) inteligente genera una avalancha de *outputs* cuya utilidad puede ser a veces cuestionable. El tiempo dirá si, a la luz de los ritmos de trabajo digitalmente densificados, la cantidad de información aumentada exponencialmente y la inflación de contenidos basados en texto, se puede hablar de un efecto rebote intelectual-digital. Una dinámica comparable ya se observa en la infraestructura digital: aunque

---

inminente la superación de este límite planetario, es decir, el llamado *green water planetary boundary* (cf. Wang-Erlandsson et al., 2022).



chips y tecnologías de almacenamiento más potentes reducen el consumo de energía por unidad de cálculo, los volúmenes de datos en crecimiento exponencial, las crecientes demandas de cargas de cálculo basadas en IA, así como las mayores necesidades de almacenamiento en el marco de la transición energética – por ejemplo, para equilibrar las inyecciones fluctuantes de energías renovables en la red eléctrica – están aumentando significativamente el consumo total de energía de los centros de datos (*hyperscalers*, *colocation*, pero también *edge datacenters*). El resultado es una fragmentación estructural en la que la eficiencia ya no se concibe como una categoría liberadora, sino como una categoría de control en un mundo de fragilidad cognitiva y disyunciones estructurales reales.

Como se ha podido demostrar, esta retroalimentación de los aumentos de eficiencia sobre el consumo total, conocida como efecto rebote, revela no en menor medida la incompletitud de la racionalidad técnica, ya que la mejora de un subsistema no genera necesariamente armonía sistémica, sino a veces fricción sistémica. La paradoja se vuelve particularmente clara donde la tecnología misma se utiliza para compensar las consecuencias de la tecnología: los *smart homes* deben ahorrar energía de forma inteligente, pero necesitan flujos de datos permanentes (IoT), *backends* de alto consumo energético y soluciones de baterías que requieren muchos recursos. Es la génesis de un nuevo nivel de complejidad, aceleración y – posiblemente – *alienación*. La

rueda de hámster de la solución de alta tecnología consiste, pues, en que los aumentos de eficiencia no rompen con el paradigma de eficiencia empleado, sino que lo perpetúan y lo producen: cada nuevo problema que la eficiencia alguna vez causó, debe ser resuelto con nuevas eficiencias; el *Uròboros* en una dependencia auto-intensificada. Como es fácil de reconocer, los aumentos de eficiencia, en realidad, no siempre conducen a un alivio, sino a nuevas formas de sobre-exigencia (cf. Gillingham et al., 2016; Vivanco et al., 2022); y esto significa, a la luz del cambio climático, que la eficiencia como categoría de la acción sostenible no es neutral en sus efectos, sino que conlleva efectos de retroalimentación que deben ser reflexionados tanto normativa como estructuralmente.

## **La adaptación como “nueva honestidad” a la luz del rebote**

La eficiencia no es una técnica neutral en cuanto a valores, sino un dispositivo de racionalidad que ha crecido cultural e ideohistóricamente, que privilegia ciertas relaciones con el mundo – controlar, optimizar, disciplinar – y las traduce en órdenes prácticos. En la medida y según esta traducción, sin embargo, se marginan otras formas de la relación hombre-naturaleza: el cuidado intercultural, la tolerancia intergeneracional de la continencia, así como la inserción en contextos mundiales no lineales. Es en este punto donde la idea de la *adaptación* – descrita por el IPCC (2023) como el proceso de ajuste al clima actual o esperado y sus

efectos – gana su agudeza ético-realpolítica: el efecto rebote revela que la *mitigación* a través de aumentos de eficiencia (es decir, tanto soluciones abiertas a la tecnología como mecanismos optimizados de evitación y reducción) *por sí sola* sigue siendo estructuralmente incompleta. De ello se sigue, con una doble necesidad, por un lado, el *reconocimiento de la incompletitud* y, por otro, el *fortalecimiento de las capacidades adaptativas*. Por lo tanto, es necesario, por un lado, reconocer que política y éticamente habrá que desprenderse de la ilusión de una mitigación exhaustiva, ya que las consecuencias negativas del cambio climático seguirán siendo palpables, incluso en los mejores escenarios de eficiencia. Por otro lado, parece necesario, en lugar de apostar *prima facie* por soluciones eficientes y *soluciones tecnológicas*, orientar la dimensión configuradora de la acción sostenible mucho más hacia infraestructuras resilientes, sistemas de seguridad social y prácticas culturalmente robustas. Filosóficamente hablando, esto requiere sin embargo una revisión de la relación medio-fin: no es la eficiencia de los medios lo que legitima la acción, sino la resiliencia del fin. Pues si las pérdidas son *inevitables* dado el curso actual de la relación con el mundo, entonces el *nomos del presente* debe orientarse a la permanencia y resiliencia de la *vida humana auténtica en la Tierra*.

A la “nueva honestidad” le está inscrita una cierta brutalidad pesimista, que sin embargo no debe ser malinterpretada como una capitulación deficitaria o una

estagnación pasiva, sino que permanece orientada a la configuración de un futuro resistente. La carga de consecuencias político-normativas que de ello se deriva abre al mismo tiempo nuevos márgenes de acción y ofrece un cierto *claro* – y con ello libertad – en la maraña de los debates climáticos actuales: junto a una *política de adaptación*, que institucionalice la adaptación socio-ecológica en la planificación urbana, los sistemas alimentarios, el suministro de energía, etc., bajo la premisa de que ocurrirán *shocks* (cf. *irreversible damages*<sup>31</sup>), es necesario crear métricas de evaluación capaces de absorber pérdidas de función y formar nuevos estados de equilibrio. En tercer lugar, una nueva *ética de la indulgencia* debe ser capaz de realizar recodificaciones conceptuales, pues la adaptación implica una actitud que cuenta con el fracaso de medidas (de eficiencia) individuales, sin estigmatizarlo como un fracaso moral o culpable. Esta indulgencia – ciertamente delicada – hacia lo (probablemente ya) inevitable no representa, sin embargo, una resignación, sino una nueva variante de la razón previsor.

---

<sup>31</sup> El IPCC entiende por daños irreversibles (*irreversible damages*) aquellos que ocurren en escalas de tiempo que exceden los horizontes de planificación humana y que, de facto, no pueden revertirse incluso si se reduce la intensidad de las presiones. Como ejemplos pueden citarse el derretimiento de grandes capas de hielo, la extinción de especies o la pérdida de ecosistemas cuya funcionalidad queda permanentemente afectada tras superar el respectivo *punto de inflexión*.

Llegados a este punto, el concepto de “indulgencia” (*Nachsicht/Milde*) exige una delimitación conceptual rigurosa para no ser confundido con una resignación pasiva o una mera estrategia discursiva. En el marco de esta propuesta, la indulgencia se conceptualiza como una categoría ético-política y una heurística de la acción socioecológica. Frente al mandato implacable y disciplinario de la optimización continua, la indulgencia representa la aceptación política y consciente de la vulnerabilidad y de los límites materiales del sistema Tierra. No se trata de una claudicación moral, sino de la decisión ética de dejar de forzar los ecosistemas mediante soluciones tecno-eficientes que, paradójicamente, aceleran la degradación. Es una disposición política que permite integrar la pérdida, el daño y la adaptación, priorizando la estabilidad a largo plazo (resiliencia) por encima de la maximización del rendimiento a corto plazo.

Entendida más de cerca, esta idea, concebida como un contraproyecto, ofrece la oportunidad de alejarse de una autodescripción deficitaria del sistema social hacia un ethos político de sostenibilidad y de vida ciertamente sobrio, pero orientado al futuro. De la crítica genealógica y analítica de la lógica de la eficiencia en su génesis semántico-conceptual se ha desprendido que, mientras la bondad de la acción se mida por la productividad de sus relaciones mediales, la política necesariamente errará el horizonte normativo y, por tanto, real de la problemática del cambio climático. Una revisión de

la misma, que ponga en primer plano la resiliencia del fin, reconocerá entonces que las pérdidas serán probablemente inevitables; precisamente por eso, un nuevo nomos del presente debe orientarse hacia la permanencia y la capacidad de resistencia de la vida humana. Resiliencia significa, por consiguiente, robustez en medio de la pérdida inevitable. Es la capacidad de un sistema no solo para sobrevivir a los *shocks*, sino para reorganizarse a través de ellos; implica una pluralización del establecimiento de fines y, por ende, de la consecución de los medios, ya que las dependencias de trayectoria se vuelven más plurales. Todo esto quiere decir: no es la dosificación de los medios, sino el mantener abiertas múltiples y culturalmente distintivas vías de adaptación y, por tanto, de supervivencia, lo que gana en importancia. Además de una política de resiliencia<sup>32</sup> previsor, especialmente el concepto de “suficiencia” – entendido como una moderación normativa en el consumo de recursos – abre una directriz central, aunque en parte controvertida, para los debates sobre política climática y ambiental. Suficiencia (cf. Held et al., 2024; Ott et al., 2007) no significa renuncia en sentido

---

<sup>32</sup> Una *política de resiliencia* significaría entonces, por ejemplo, que la construcción de infraestructuras aparentemente secundarias, como el suministro descentralizado de energía y agua, múltiples vías de transporte o redes de comunicación aparentemente redundantes, se considere funcional; tecnológicamente, significa que las innovaciones técnicas no apunten únicamente al aumento de la eficiencia, sino también a la tolerancia a fallos, la reparabilidad, la durabilidad y la adaptabilidad modular. Esto es más que economía circular; es la preservación frente a lo imperfecto e inminente.

estricto, sino la autolimitación consciente de las necesidades y el consumo a un nivel ecológica y socialmente sostenible, con lo que los criterios normativos se sitúan en el centro de la controversia; son debates sobre la cuestión, largamente reprimida, de la redefinición de *qué* se considera en absoluto un objetivo sensato y *en qué medida*. Es la revisión probablemente más radical de la relación medio-fin: la política de suficiencia no solo exige decisiones políticas valientes – ya sea la limitación de bienes y soluciones de movilidad intensivos en energía, impuestos progresivos sobre los recursos, o límites claros de renuncia y consumo. Con ella van asociadas, socio-estructural y – culturalmente, no solo huellas de carbono (*Carbon Footprints*) o hídricas (*Water Footprints*) claramente medibles<sup>33</sup>, sino también la emergencia de nuevos estilos de vida que apuestan conscientemente por relaciones de propiedad menos individualistas y por acuerdos de uso compartido, a menudo en combinación con discursos de desaceleración y atención plena (*mindfulness*).

---

<sup>33</sup> Por *huella de carbono* (*Carbon Footprint*) se entiende la cantidad total de emisiones de gases de efecto invernadero, causadas directa e indirectamente por individuos, organizaciones, productos o actividades, expresada en equivalentes de CO<sub>2</sub>. La *huella hídrica* (*Water Footprint*), por su parte, designa el volumen total de recursos de agua dulce necesarios para producir un producto o prestar un servicio, incluyendo el consumo de agua directo e indirecto a lo largo de toda la cadena de valor.

Este no es el lugar para realizar una evaluación normativa, una investigación lógico-crítica o un análisis de la viabilidad práctica global de estas narrativas; lo importante parece ser señalar que *en estos discursos* se abre un *espacio de resonancia cultural* que establece una semántica diferente de bienestar, tiempo, salud y bien común. Si se lleva la idea de la *desaceleración* (cf. Princen, 2005; Schneidewind y Zahrt, 2013) un paso más allá en su dimensión ético-progresiva, se llega, en tercer lugar, al paradigma de la *desoptimización*, la ralentización consciente; esta puede interpretarse tanto de forma sostenible-analógica como digital-específica de la IA. Pues mientras que la resiliencia y la suficiencia todavía apuntan a la estabilización y la limitación, la idea de la desoptimización asesta un golpe ético crítico a la propia heurística de la eficiencia. Pues desoptimización significa entonces diseñar sistemas para que no funcionen de manera *máxima*, sino *atenta*; piénsese en cadenas de suministro deliberadamente ralentizadas y, a cambio, más locales, procesos de producción más robustos en lugar de optimizados para el *just-in-time*, o aplicaciones digitales que evolucionan más lentamente. La idea decisiva de los tres conceptos – resiliencia, suficiencia, desoptimización – radica en que obtienen su *urgencia* propia (y por tanto, social-normativa) no en menor medida de la comprensión de que la mitigación seguirá siendo incompleta. Los esfuerzos de protección climática deben, por esta razón, madurar hacia una *política de adaptación*, es decir, deben reaccionar a nivel



infraestructural (resiliencia de los asentamientos, agricultura adaptativa,...), a nivel socioeconómico (redes de seguridad en forma de ayudas y subvenciones,...) así como a nivel de política global (*Loss and Damage*, marcos legales para refugiados climáticos,...). Aunque compleja, controvertida y multifacética en los detalles, con todas estas ideas se asocia finalmente una *ética de la indulgencia*; es decir, la capacidad de no denunciar, descartar y devaluar moralmente el fracaso de medidas individuales o incluso colectivas-agregadas, sino de aceptarlo como parte de una estrategia de adaptación política-realista. Esta actitud puede ser entendida a menudo como una defensa frente a la exigencia de una mitigación drástica o una responsabilidad global e histórico-justa. Por ello, es tanto más importante comunicar la indulgencia no como un discurso de evasión, sino como una perspectiva de configuración conjunta que incluye al mismo tiempo la responsabilidad por los daños y el apoyo a los grupos y comunidades más vulnerables. La crítica al dispositivo de la eficiencia nos obliga, en última instancia, a replantear la responsabilidad intergeneracional. La crisis ecológica no admite dilaciones justificadas por la promesa de futuras innovaciones tecnológicas que supuestamente corregirán las ineficiencias actuales. En consecuencia, la responsabilidad hacia las generaciones venideras no radica en legarles un mundo ilusoriamente “perfectamente optimizado”, sino en asumir desde el presente los costos sociopolíticos de la suficiencia y la resiliencia. El imperativo ético se desplaza así

de la maximización tecno-económica hacia la preservación estricta de los límites planetarios, asumiendo nuestra ineludible responsabilidad histórica como causantes de su transgresión.

## **Conclusión y perspectivas**

Aunque una “nueva honestidad” así orientada lleva consigo una cierta brutalidad pesimista fundamental al reconocer ciertas pérdidas y trastornos como inevitables, no es un llamamiento a un nihilismo deficitario, sino una forma de *humanismo práctico-robusto* (cf. Sailer, 2025): una política que no solo se orienta a la mejora efectiva de los medios, sino a la resiliencia del fin, es decir, a la permanencia de la vida humana auténtica con dignidad. Aquí, la crítica genealógica de la lógica de la eficiencia, a través del ejemplo del efecto rebote, muestra que la orientación hacia la maximización de los medios en cuestiones de política climática conduce a veces a callejones sin salida no intencionados. Y son los trastornos inminentes del cambio climático mismo, a lo largo de los límites planetarios y la superación de los *puntos de inflexión* climáticos críticos, los que forzarán una revisión de la estructura de relación medio-fin predominante. No debería ser la eficiencia la que – normativamente hablando – legitime fundamentalmente la acción colectiva, sino la *resiliencia del fin*, es decir, *teleológicamente* hablando, la *Eudaimonia* de la vida humana con dignidad, orientada a una existencia duradera.

Esto requiere, sin embargo, esa “nueva honestidad” que de ninguna manera abandona la mitigación – ni política, ni tecnológica, ni económico-financieramente– pero que reconoce su incompletitud y, por tanto, apuesta simultáneamente por la adaptación, la indulgencia y una configuración proactiva de las vulnerabilidades como categorías ampliadas de la acción sostenible. En esto no hay una capitulación, sino una rearticulación de orden superior de esa misma libertad que permite la configuración de un futuro resistente, con la conciencia crítica de pérdidas, tensiones y líneas de fractura inevitables, y que *también* permite intervenciones en la naturaleza y, *precisamente en eso*, actúa de manera responsable. Tales intervenciones en la naturaleza son, por supuesto, *disciplinadas*, solo que *naturalmente* disciplinadas.

## Referencias

- Angehrn, E. (2023). Hermeneutik des Lebens. En S. Hüsch y O. Victor (Eds.), *Das Konzept Leben in der Geschichte der Philosophie. Le concept de vie dans l'histoire de la philosophie*. Ergon.
- Berkhout, P. H. G., Muskens, J. C. y Velthuijsen, J. W. (2000). Defining the rebound effect. *Energy Policy*, 28(6-7), 425-432.
- Böhme, H. (Ed.). (1988). *Kulturgeschichte des Wassers*. Suhrkamp.

- Brockway, P. E., Sorrell, S., Semieniuk, G., Heun, M. K. y Court, V. (2021). Energy efficiency and economy-wide rebound effects: A review of the evidence and its implications. *Renewable and Sustainable Energy Reviews*, 141, 110781.
- Crutzen, P. J. y Stoermer, E. F. (2006). The Anthropocene. En E. Ehlers y T. Krafft (Eds.), *Earth System Science in the Anthropocene* (pp. 13-18). Springer.
- Descartes, R. (2011). *Discours de la méthode*. Meiner.
- Foucault, M. (2004). *Sécurité, territoire, population. Cours au Collège de France, 1977-1978*. Gallimard/Seuil.
- Fronzel, M., Ritter, N. y Vance, C. (2012). Heterogeneity in the rebound effect: A meta-analysis of studies on household energy demand. *Energy Economics*, 34(5), 1691-1701.
- Gadamer, H.-G. (1999). *Der Anfang des Wissens*. Reclam.
- Galvin, R. (2015). The rebound effect, energy efficiency and the Jevons Paradox: A case for breaking out of the straitjacket of neoclassical economics. *Energy*, 82, 634-643.
- Gehlen, A. (1976). *Die Seele im technischen Zeitalter. Sozialpsychologische Probleme in der industriellen Gesellschaft*. Rowohlt.
- Gillingham, K., Rapson, D. y Wagner, G. (2016). The Rebound Effect and Energy Efficiency Policy. *Review of Environmental Economics and Policy*, 10(1), 68-88.

- Held, B., Rodenhäuser, D., Vetter, H. y Diefenbacher, H. (Eds.). (2024). *Suffizienzpolitik: Möglichkeiten und Grenzen in einer sozial-ökologischen Marktwirtschaft* (Die Wirtschaft der Gesellschaft, N.º 9). Metropolis-Verlag.
- Herring, H. y Roy, R. (2007). *Technological Innovation, Energy Efficient Design and the Rebound Effect*. Elgar.
- IPCC. (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report*. <https://www.ipcc.ch/report/ar6/wg2/>
- IPCC. (2023). *Climate Change 2023: Synthesis Report*.
- Jevons, W. S. (1865). *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines*. Macmillan.
- Jonas, H. (1984). *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*. Suhrkamp.
- Kerkmann, J. y Sailer, D. M. (en prensa). *Verlaufsrichtungen des politischen Denkens. Geschichtsphilosophische Theorien zwischen Verfall und Fortschritt*. Nomos.
- Luhmann, N. (1998). *Die Gesellschaft der Gesellschaft* (Vols. 1-2). Suhrkamp.
- Luhmann, N. (2014). *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität*. UVK.

- Ott, K. y Voget, L. (2007). *Suffizienz: Umweltethik und Lebensstilfragen* (Vordenken – Ökologie und Gesellschaft, N.º 2). Heinrich-Böll-Stiftung.
- Princen, T. (2005). *The Logic of Sufficiency*. MIT Press.
- Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., III, Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., . . . Foley, J. A. (2009). A safe operating space for humanity. *Nature*, 461, 472-475. <https://doi.org/10.1038/461472a>
- Sailer, D. M. (2022). *Die Disziplinierung des Wassers. Eine kultur- und ideengeschichtliche Analyse* (Historisch-anthropologische Studien, Bd. 28). Lang.
- Sailer, D. M. (2023). Klimaethik und Naturdisziplinierung. Der Grenzbegriff als klimaethische Grundnorm. En Schweizerische Philosophische Gesellschaft (Ed.), *Studia Philosophica. Schweizerische Zeitschrift für Philosophie. Ökologie und Ethik* (Vol. 81). Schwabe.
- Sailer, D. M. (2025). Robuster Humanismus als Politischer Realismus. En J. Kerkmann, F. Herkert, D. M. Sailer y R. Voigt (Eds.), *Politischer Realismus II: Zur ideengeschichtlichen und geopolitischen Entfaltung in der Moderne* (pp. 209-248). Nomos. <https://doi.org/10.5771/9783748947400-209>

- Sailer-Wlasits, P. (2020). *Uneigentlichkeit. Philosophische Besichtigungen zwischen Metapher, Zeugenschaft und Wahrsprechen*. Königshausen & Neumann.
- Santarius, T. (2016). *Der Rebound-Effekt: Ökonomische, psychische und soziale Herausforderungen für die Entkopplung von Wirtschaftswachstum und Umweltverbrauch*. Metropolis.
- Santarius, T., Walnum, H. J. y Aall, C. (Eds.). (2016). *Rethinking Climate and Energy Policy: New Perspectives on the Rebound Effect*. Springer.
- Schipper, L. y Grubb, M. (1993). On the rebound? Feedback between energy intensities and energy uses in IEA countries. *Energy Policy*, 21(5), 493-518.
- Schneidewind, U. y Zahrnt, A. (2013). *Damit gutes Leben einfacher wird: Perspektiven einer Suffizienzpolitik*. oekom.
- Sorrell, S. (2007). *The Rebound Effect: An Assessment of the Evidence for Economy-wide Energy Savings from Improved Energy Efficiency*. UK Energy Research Centre.
- Sorrell, S. y Dimitropoulos, J. (2008). The rebound effect: Microeconomic definitions, limitations and extensions. *Ecological Economics*, 65(3), 636-649.
- Steffen, W., Rockström, J., Noone, K., Persson, Å., Chapin, F. S., III, Lambin, E., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S.,

Snyder, P. K., Costanza, R., Svedin, U., . . . Foley, J. (2009). Planetary boundaries. Exploring the safe operating space for humanity. *Ecology and Society*, 14(2), artículo 32. <https://www.ecologyandsociety.org/vol14/iss2/art32/>

- Vivanco, D. F., Echeverria, J. A. y Garmendia, E. (2022). A systematic literature review on the circular economy rebound. *Journal of Cleaner Production*, 373, 133816.
- Wallimann-Helmer, I., Meyer, L., Mintz-Woo, K., Schinko, T. y Serdeczny, O. (2019). The Ethical Challenges in the Context of Climate Loss and Damage. En R. Mechler, L. M. Bouwer, T. Schinko, S. Surminski y J. Linnerooth-Bayer (Eds.), *Loss and Damage from Climate Change: Concepts, Methods and Policy Options* (pp. 39-62). Springer.
- Wang-Erlandsson, L., Tobian, A., van der Ent, R. J., Fetzer, I., te Wierik, S., Porkka, M., Staal, A., Jaramillo, F., Dahlmann, H., Singh, C., Greve, P., Gerten, D., Keys, P. W., Gleeson, T., Cornell, S. E., Steffen, W., Bai, X. y Rockström, J. (2022). A planetary boundary for green water. *Nature Reviews Earth & Environment*, 3(6), 380-392. <https://doi.org/10.1038/s43017-022-00287-8>



# Pensamiento político contemporáneo



# **Presentación de la sección “Pensamiento político contemporáneo”**

**Juan Esteban Echeverry Murillo<sup>34</sup>**

Esta sección del libro reúne doce capítulos que examinan, desde tradiciones teóricas diversas, las transformaciones del poder, de la legitimidad y de la vida democrática en el siglo XXI. Si algo comparten estos trabajos es la convicción de que el pensamiento político no puede limitarse hoy a administrar categorías heredadas: debe someterlas a prueba frente a fenómenos como la revolución algorítmica, la desigualdad estructural, la emocionalización del discurso público y la recomposición de los órdenes regionales. El recorrido propuesto va de los fundamentos a sus disputas, y de estas a sus concreciones históricas.

La sección se abre con "Ser, datos y poder. Una lectura crítica del sujeto datificable en la era del control algorítmico",

---

<sup>34</sup> Fundación Info Kratos. <https://orcid.org/0009-0007-0757-2369>

que sitúa desde el inicio el problema que atravesará buena parte de los capítulos siguientes: convertir la existencia humana en dato constituye una operación de poder que transforma la estructura misma de la subjetividad. El sujeto datificable, sostienen los autores, es la figura central de la tecnopolítica actual, producido desde adentro por arquitecturas algorítmicas que extraen, modelizan y anticipan la conducta humana antes de cualquier deliberación. A partir de una genealogía crítica del sujeto moderno y de un análisis de la racionalidad de la gubernamentalidad algorítmica y el smart power, el capítulo examina la crisis ontológica del ser bajo condiciones de datificación total, para concluir que gobernar mediante datos representa una mutación en la relación entre poder, verdad y experiencia humana que compromete las condiciones mismas de la acción política autónoma.

El segundo capítulo, "Humanismo cívico: un legado de la filosofía política clásica para el mundo contemporáneo", por Carlos Andrés Gómez Rodas y Elvira Bedoya Argüelles, recupera la filosofía política de Aristóteles, en especial su concepción teleológica de la naturaleza humana, para reivindicar, frente al individualismo liberal moderno, la comunidad política como ámbito natural de realización personal y moral. La centralidad de la virtud, de la amistad cívica y del bien común, junto con el papel formativo de la familia y la escuela, configura una alternativa ética y política que articula libertad, verdad y comunidad desde una

ontología realista y finalista. Frente a la crisis del sujeto que abre la sección, este capítulo ofrece así una respuesta desde los fundamentos clásicos.

El tercer capítulo, "Desigualdad neoliberal, neofascismo, tecnofeudalismo y necropolítica. Claves para disputar la democracia", por Julio Peña y Lillo E., ofrece el contrapunto crítico. En él se sostiene que la desigualdad neoliberal no es un fenómeno económico neutral sino una decisión política deliberada que produce exclusión estructural y vulnerabilidad social, y se analizan el neofascismo, la necropolítica y el tecnofeudalismo como sus mutaciones extremas. Frente a estas formas autoritarias de gestión social, el capítulo defiende la urgencia de recuperar el Estado como garante de derechos y condición de sociedades más justas, inclusivas y democráticas.

Los dos capítulos siguientes profundizan en el escenario propiamente tecnológico ya anunciado al inicio de la sección. "Inteligencia artificial y tecnoceno: genealogía, control y dilemas éticos de la revolución algorítmica", por Ana Sedeño-Valdellos y Xavier Brito Alvarado, reconstruye, desde la teoría crítica y con apoyo en autores como Bostrom, Sadin, Floridi, Žižek y Land, el recorrido histórico y conceptual de la inteligencia artificial, para examinar luego sus efectos normalizadores, su capacidad de control y los riesgos existenciales y narrativas alternativas (como el aceleracionismo) que la acompañan. Por su parte,

"Gubernamentalidad algorítmica: líneas de fuga para una crítica del poder digital contemporáneo", por Ivonne Jiménez Jiménez, examina la transformación del poder hacia un régimen predictivo que administra la vida mediante el seguimiento continuo de datos y, recurriendo al concepto deleuziano de líneas de fuga, propone una cartografía analítica articulada en cuatro ejes (subjetivación biopolítica, vigilancia y control, opacidad y manipulación, desigualdad estructural y resistencias) que muestra cómo la maquinaria algorítmica captura y reestratifica incesantemente las fisuras del sistema.

Ahora bien, si el poder digital reconfigura la dominación, también transforma las posibilidades de la ciudadanía. "Mecanismos de participación ciudadana en la sociedad digital", por María Fernanda Idrobo Flores y Boris Raúl Ochoa Jiménez, analiza cómo las tecnologías de la información han ampliado los escenarios de deliberación pública, incidencia política y control social, promoviendo una participación más abierta, inmediata y descentralizada, sin ocultar los obstáculos que persisten: la brecha digital, la falta de alfabetización tecnológica y la escasa institucionalización de los espacios digitales. La participación en la era digital aparece así como una oportunidad para fortalecer la democracia deliberativa, condicionada al acceso equitativo, la transparencia y la protección de los derechos fundamentales en línea.

El séptimo capítulo desplaza el análisis hacia un caso latinoamericano de primera magnitud. "Emociones y movilización política en el sexenio de AMLO (2018-2024). La reconfiguración de la política mexicana en el contexto de la Cuarta Transformación", por Ekhiñe Graell Larreta, examina el papel central de los afectos en la articulación de identidades políticas, mostrando cómo las narrativas de la esperanza, el amor, la reivindicación ética y la simbología religiosa construyeron un horizonte moral diferenciado entre "pueblo bueno" y adversarios, y cómo la comunicación presidencial reforzó la polarización afectiva desplazando los debates programáticos hacia significados morales. La emocionalización se revela, en este estudio, como estrategia de poder y como proceso estructural de la política contemporánea.

En "Itinerarios del pensamiento político y las hegemonías en la configuración del sujeto social", por Pablo de Miguel, la sección recobra la mirada de conjunto: desde la polis griega hasta las formas comunales contemporáneas, el capítulo muestra cómo continuidades y rupturas han configurado al sujeto social en relación con distintas hegemonías, y cómo la posverdad, las fake news y la desconfianza en la ciencia constituyen hoy desafíos centrales para la democracia, proyectando un modelo plural y emancipador del sujeto social para el siglo XXI.

Los dos capítulos siguientes vuelven sobre los fundamentos normativos de la democracia desde tradiciones distintas y complementarias. "Cristianismo y democracia moderna: la influencia de la doctrina social de la Iglesia en la construcción de instituciones democráticas", por Javier Fattah y Carolina Lagos, expone, a partir de la documentación eclesiástica y en particular de la encíclica *Rerum Novarum*, la contribución del pensamiento social cristiano al compromiso democrático con el bien común y la dignidad de la persona, transitando de la inicial tensión entre cristianismo y modernidad hacia una relación imprescindible entre ambos órdenes de valores. A su vez, "La crítica rapsódica de Rawls al capitalismo: un fundamento antropológico para la democracia de propietarios", por Javier Enrique Castillo Vallez, sostiene que la crítica rawlsiana al capitalismo de bienestar, lejos de ser fragmentaria, revela un fundamento antropológico coherente que encuentra en la democracia de propietarios la realización institucional más fiel de la justicia como equidad, al asegurar las condiciones materiales para el desarrollo de los dos poderes morales de la persona.

El undécimo capítulo, "Gobernanza ética y psicopolítica digital: razón algorítmica, poder y circulación social del conocimiento", por Paula Andrea Mora Pedreros y Manuel José Santiago Pajajoy, opera como síntesis normativa de las preocupaciones de la sección: distingue entre la psicopolítica digital, lógica de subjetivación que opera mediante personalización, optimización y economía de la

atención, y la gobernanza ética, marco normativo e institucional capaz de orientar, limitar y auditar esa racionalidad algorítmica en nombre de la dignidad, la responsabilidad y la deliberación democrática, articulando para ello los aportes de Byung-Chul Han, Adela Cortina y Daniel Innerarity.

La sección cierra con "Trade, Security, and Selective Alignment: Process Tracing the Peru-United States Trade Promotion Agreement and the Reorientation of Peruvian Foreign Policy", por Anthony Rolando Medina Rivas Plata y Aaron Ubaldo del Carpio Retamozo, que traslada el examen del poder contemporáneo a la escala internacional. Mediante la técnica del process tracing, el capítulo reconstruye cómo el acuerdo comercial entre Perú y Estados Unidos funcionó, más allá de su dimensión económica, como mecanismo de alineamiento estratégico que consolidó el modelo peruano de apertura y reorientó su política exterior en un orden andino ideológicamente disputado. Este cierre recuerda que las ideas, las instituciones y los afectos examinados a lo largo de la sección se despliegan siempre dentro de configuraciones geopolíticas concretas, y que pensar políticamente el siglo XXI exige atender, a la vez, a los fundamentos, a los poderes que los desafían y a los órdenes que los enmarcan.



# Ser, datos y poder. Una lectura crítica del sujeto datificable en la era del control algorítmico

*Being, Data, and Power. A Critical Reading of the Datafied Subject in the Age of Algorithmic Control*

**Santiago Andrés Ullauri Betancourt<sup>35</sup>, Diego Ortiz-Jaramillo<sup>36</sup>**

## Resumen

Convertir la existencia humana en dato constituye una operación de poder que transforma la estructura de la subjetividad y reconfigura los modos en que el poder contemporáneo actúa, conoce y gobierna. Este capítulo

---

<sup>35</sup> Instituto para el Desarrollo de la Cultura y Sociedad - Universidad Hemisferios. <https://orcid.org/0000-0003-0858-3178>

<sup>36</sup> Media & Innovation Lab - Universidad Hemisferios. <https://orcid.org/0009-0007-3460-5824>

sostiene que el sujeto datificable representa la figura central de la tecnopolítica actual, producido desde adentro por arquitecturas algorítmicas que extraen, modelizan y anticipan la conducta humana antes de cualquier deliberación. El argumento recorre una genealogía crítica del sujeto moderno, analiza la racionalidad de la gubernamentalidad algorítmica y el *smart power*, examina la crisis ontológica del ser bajo condiciones de datificación total y extrae las consecuencias político-democráticas de ese proceso. La conclusión es que gobernar mediante datos representa una mutación en la relación entre poder, verdad y experiencia humana que compromete las condiciones mismas de la acción política autónoma.

**Palabras clave:** Datificación; psicopolítica; control algorítmico; *smart power*; subjetividad; tecnopolítica.

## **Abstract**

Converting human existence into data is not a neutral act of registration. It is an operation of power that transforms the very structure of subjectivity and reconfigures the ways in which contemporary power acts, knows, and governs. This chapter argues that the datafied subject constitutes the central figure of contemporary technopolitics, produced from within by algorithmic architectures that extract, model, and anticipate its conduct prior to any deliberation. The argument traces a critical genealogy of the modern subject, analyzes the

rationality of algorithmic governmentality and smart power, examines the ontological crisis of being under conditions of total datafication, and draws out the political-democratic consequences of that process. The conclusion is that governing through data represents a mutation in the relationship between power, truth, and human experience that compromises the very conditions of autonomous political action.

**Keywords:** Datafication; psychopolitics; algorithmic control; smart power; subjectivity; technopolitics.

## **Introducción**

La pregunta por el sujeto en la era digital no nace con las plataformas ni con los algoritmos de recomendación. Hunde sus raíces en una transformación más profunda y menos advertida del capitalismo contemporáneo, el descubrimiento de que la conducta humana puede ser convertida en dato, el dato en predicción y la predicción en un recurso económica y políticamente explotable. Antes de que existiera un clic, ya operaba la convicción de que el ser humano podía ser conocido, clasificado y gobernado a través de su comportamiento observable, y esa convicción tenía tras de sí una larga historia que va desde las técnicas de registro estadístico del siglo XIX hasta las formas más sofisticadas de vigilancia industrial y gubernamental. Lo que el capitalismo de plataformas ha hecho es tomar esa convicción antigua y

dotarla de una infraestructura técnica sin precedentes, capaz de operar a una escala y con una granularidad que las formas previas de clasificación y registro no podían ni imaginar.

Abordar este fenómeno exige resistir la tentación de tratarlo como un problema meramente tecnológico. La datificación no es una operación que acontece sobre una subjetividad preexistente e intacta, un procedimiento neutro que deja al sujeto tal como lo encontró. Participa activamente en la producción del sujeto que dice conocer, y esa circularidad entre producción y conocimiento es uno de los rasgos más inquietantes del régimen tecnopolítico contemporáneo. Cuando las plataformas extraen trazas de comportamiento, construyen perfiles de probabilidad y diseñan entornos que anticipan la conducta del usuario antes de que este delibere, contribuyen a producir una forma específica de subjetividad con sus propias estructuras de deseo y sus propias cegueras constitutivas. Reducir este proceso a una cuestión de privacidad o de regulación tecnológica equivale a tratar un síntoma sin interrogar la enfermedad que lo produce.

La pregunta que orienta este capítulo puede formularse así. ¿De qué manera la datificación sistemática de la existencia humana transforma la estructura de la subjetividad, las formas de dominación y las condiciones de posibilidad de la acción política autónoma? No es una pregunta empírica sobre los efectos observables de la

tecnología, es una pregunta filosófico-política que interroga la relación entre poder y subjetividad, entre verdad y gobierno, entre libertad y control, y esa distinción importa porque un problema meramente técnico admite solución regulatoria, mientras que un problema filosófico-político exige también una reconstrucción de las categorías con que pensamos al sujeto y al poder.

La tesis central sostiene que la datificación del sujeto constituye una mutación onto-política, una transformación que afecta simultáneamente al ser del sujeto y a las relaciones de poder en que ese sujeto se constituye. El sujeto datificable es producido desde adentro por las arquitecturas de extracción, modelización y anticipación conductual propias del capitalismo de plataformas, y esa producción opera mediante una forma de poder que puede designarse como *smart power*, un ejercicio que no prohíbe ni coacciona, que seduce, personaliza, perfila y actúa preventivamente sobre las posibilidades de la conducta antes de que el sujeto delibere. La eficacia de este poder reside en su capacidad de producir probabilidades, de hacer más probable aquello que sirve a sus intereses sin que el sujeto perciba restricción alguna sobre su libertad. El problema filosófico y político de fondo es que ese poder actúa precisamente en el lugar donde el sujeto cree ser más libre.

El capítulo avanza en seis momentos. El primero reconstruye genealógicamente el tránsito del sujeto

disciplinario al sujeto de rendimiento datificado. El segundo examina el núcleo conceptual de la psicopolítica y sus límites teóricos internos. El tercero analiza la lógica de las plataformas y de la gubernamentalidad algorítmica como formas de conducción flexible de la conducta. El cuarto interroga el problema ontológico del ser bajo condiciones de datificación total, con especial atención a la pérdida de opacidad, de interioridad y de capacidad de juicio. El quinto extrae las consecuencias político-democráticas de estas transformaciones. El sexto cierra con un balance del rendimiento teórico del capítulo y con las preguntas que permanecen abiertas para la investigación futura.

## **Genealogía del sujeto datificable**

### *Del sujeto disciplinario al sujeto de rendimiento*

En *Vigilar y castigar*, Foucault (1976) mostró que el sujeto moderno no preexiste a las relaciones de poder que lo atraviesan, pues emerge de ellas. Las instituciones disciplinarias producen al sujeto mediante tecnologías de normalización, jerarquización y examen que hacen del cuerpo el campo privilegiado de inscripción del poder. La disciplina vuelve al cuerpo dócil y útil, lo distribuye en el espacio, lo somete a la mirada continua y le asigna un lugar en la grilla normativa que separa lo normal de lo patológico. La subjetividad moderna surge en ese proceso de formación, y el yo capaz de reflexionar sobre sí mismo y de examinarse es

también el yo que ha sido producido por las técnicas de poder que lo hacen legible, evaluable y administrable. La conciencia de sí no es una conquista del sujeto frente al poder, es, en buena medida, un efecto de ese poder sobre el sujeto, lo que revela la dimensión productiva y no meramente represiva de las relaciones de dominación modernas.

Lo que este análisis pone en juego, más allá de su dimensión histórica específica, es un principio de alcance general. La subjetividad es siempre el resultado de prácticas históricamente situadas de producción del yo, y cada configuración histórica del poder produce el tipo de sujeto que necesita para funcionar. Cuando el capitalismo digital despliega nuevas formas de extracción de datos, de perfilamiento de la conducta y de modulación algorítmica del entorno, participa en la producción de un tipo nuevo de subjetividad, y el capitalismo de plataformas no es una excepción a esa lógica, es su realización más perfecta.

Esta comprensión adquiere mayor profundidad al articularse con el análisis foucaultiano de la racionalidad neoliberal. Foucault (2007) mostró que el neoliberalismo opera como una tecnología de gobierno que produce activamente un tipo de sujeto determinado, el *homo economicus* que aprende a comprenderse a sí mismo como empresa, que evalúa todas sus decisiones bajo la lógica de la inversión y la rentabilidad, y que convierte cada aspecto de su vida en objeto de gestión estratégica. El neoliberalismo no se

limita a gobernar sujetos ya constituidos, los produce, les enseña a pensarse como unidades de capital humano que deben administrarse con eficiencia, y esa producción gubernamental del sujeto es la condición de posibilidad histórica sobre la que el capitalismo de datos construye su operación más específica, un sujeto predispuesto a concebirse como recurso optimizable y, por lo mismo, mucho más permeable a las lógicas de perfilamiento y anticipación conductual.

En su *Posdata sobre las sociedades de control*, Deleuze (1990) señaló que las instituciones disciplinarias estaban en crisis y que emergía una nueva modalidad de poder que opera mediante modulaciones continuas, flujos y redes. En las sociedades disciplinarias, el individuo pasaba de la familia a la escuela, de la escuela a la fábrica, describiendo un itinerario de moldes sucesivos y discontinuos, mientras que en las sociedades de control los umbrales se desplazan, los controles son continuos y la distinción entre el adentro y el afuera de la institución se desvanece. El sujeto ya no es encerrado en un molde institucional definido, es modulado constantemente por variaciones imperceptibles que ajustan su conducta sin necesidad de prohibiciones explícitas. Lo que Deleuze percibía entonces como tendencia incipiente se ha convertido, con el desarrollo de las plataformas digitales y la economía de datos, en la forma dominante de gobierno de la conducta, realizando con una exactitud asombrosa aquello que el filósofo apenas alcanzaba a esbozar.



Esta transición penetra en los niveles más profundos de la producción misma de la subjetividad. Lazzarato (2014) ha argumentado que el capitalismo contemporáneo produce subjetividad como cualquier otro bien de consumo, y la distinción que propone entre sujeción social y esclavitud maquínica resulta especialmente pertinente para iluminar el funcionamiento de las plataformas. La sujeción social interpela al sujeto como individuo autónomo, consumidor, ciudadano, trabajador, reconociéndolo como sujeto y asignándole una identidad con la que puede identificarse. La esclavitud maquínica, por su parte, integra al ser humano como componente de una máquina que lo precede y lo excede, haciendo funcionar al sujeto como pieza de ensamblajes técnico-semióticos que operan por debajo del umbral de la conciencia y que no necesitan su comprensión ni su adhesión para producir efectos. Las plataformas digitales operan simultáneamente en ambos registros, y esa simultaneidad constituye uno de sus rasgos más distintivos, pues interpelan al usuario como sujeto libre que elige, personaliza y se expresa, mientras lo integran como nodo de una maquinaria de extracción que procesa sus comportamientos con una eficacia que depende precisamente de que el sujeto no advierta que está siendo procesado.

Han (2012) lleva este diagnóstico un paso más lejos al caracterizar la sociedad contemporánea como una sociedad del rendimiento en la que el poder opera mediante el imperativo de la positividad, la autooptimización y la

autoexplotación. El sujeto contemporáneo se explota voluntariamente en la carrera por el rendimiento y la productividad, y la violencia de esta forma de poder, que produce depresión y agotamiento, es invisible porque el responsable se confunde con el propio yo que persigue sus metas. A diferencia de la violencia disciplinaria, esta no necesita un agente externo para perpetuarse, funciona desde adentro del sujeto, y esa interiorización del mandato productivo la hace casi imposible de reconocer como violencia.

La transición del disciplinamiento a la psicopolítica tiene una consecuencia crucial para comprender el problema del sujeto datificable, dado que el poder neoliberal ha logrado que el sujeto interiorice la coacción como deseo propio. La autooptimización funciona al mismo tiempo como práctica de libertad y como forma de sujeción, y lo que parecía emancipación resulta ser una modalidad más profunda de dominación porque elimina la exterioridad del poder y lo instala en el corazón de la subjetividad. El sujeto neoliberal no obedece a un amo externo, obedece a la versión de sí mismo que ha interiorizado como ideal. Esta observación, que Han desarrolla con fuerza retórica considerable, requiere la precisión teórica que Ullauri Betancourt et al. (2024) aportan al señalar que el riesgo del argumento haniano es que la psicopolítica aparezca como una estructura totalizante que no deja espacio para la resistencia ni para la diferencia interna del sujeto, dado que Han tiende a presentar al sujeto como

completamente absorbido por la lógica del rendimiento, sin suficiente atención a las fisuras, los excedentes y las zonas de opacidad que persisten incluso en el interior de ese régimen. Toda teoría del poder que no reserve un lugar para la resistencia corre el riesgo de convertirse, involuntariamente, en una apología de la omnipotencia del poder que pretende criticar.

### *Autoexplotación, transparencia y psicopolítica*

La categoría central que Han despliega para pensar el poder en la era digital es la de psicopolítica. A diferencia de la biopolítica foucaultiana, que toma como objeto el cuerpo y la vida biológica de las poblaciones, la psicopolítica actúa sobre la psique, los deseos, las emociones y los afectos. Han (2014a) sostiene que el neoliberalismo digital ha encontrado en la psique su campo de extracción más rentable, pues las plataformas digitales funcionan como dispositivos que explotan la interioridad del sujeto como recurso productivo. Lo que el capitalismo industrial extraía del cuerpo mediante la disciplina, el capitalismo de datos lo extrae de la psique mediante la seducción de la interfaz, y lo hace de manera invisible porque el sujeto colabora en esa extracción al creer que se expresa cuando en realidad se produce como dato.

El poder disciplinario operaba sobre el cuerpo y necesitaba un espacio de inscripción material que podía ser localizado y confrontado. El poder psicopolítico, en cambio,

habita las arquitecturas digitales y las interfaces diseñadas para capturar la atención, de modo que el sujeto no puede señalar desde afuera el punto de aplicación del poder porque ese punto está en su interior, en la forma misma en que desea, evalúa y se relaciona consigo mismo. Han (2014b) desarrolla esta idea en *En el enjambre* al mostrar que la masa digital se fragmenta en individuos hiperconectados pero profundamente solitarios, cuya energía política se gasta en la autoexposición y la competencia por la visibilidad, produciendo ruido sin articular voz.

Han (2017) profundiza esta idea al subrayar la pérdida del respeto y la distancia que caracterizaba a la esfera pública clásica. El entorno digital fabrica una masa sin rostro que reacciona de manera inmediata y visceral ante cualquier estímulo, sin la mediación reflexiva que permite reconocer al otro como interlocutor legítimo. Lo que suele diagnosticarse como crisis de la esfera pública digital es, desde esta perspectiva, su funcionamiento esperado, pues las plataformas fueron diseñadas para la reacción, y la reacción es el modo de interacción que maximiza el *engagement* y la extracción de datos.

La transparencia es la condición cultural en que esta psicopolítica se despliega con mayor eficacia. Han (2013) muestra cómo la exigencia de transparencia, que se presenta como valor democrático, funciona en realidad como imperativo de control, produciendo un panóptico invertido en

el que los vigilados colaboran activamente en su propia vigilancia al exhibir su vida interior en las redes sociales y compartir sus preferencias con sistemas algorítmicos que los perfilarán con precisión creciente. La transparencia disuelve la opacidad que protege la singularidad del sujeto y lo convierte en superficie legible, predecible y operable. Lo que se pierde en este proceso va más allá de la privacidad jurídica, pues se pierde la alteridad, la dimensión en que el otro excede lo que ya se sabe de él, y con esa pérdida se desvanece la posibilidad del encuentro genuino con lo que no estaba ya contenido en el algoritmo de predicción.

El diagnóstico de Han es, en sus líneas fundamentales, certero, aunque opera a un nivel de abstracción que oscurece diferencias importantes. Brown (2015) ha mostrado que la racionalidad neoliberal no produce un sujeto universal, pues genera formas diferenciadas de exposición y vulnerabilidad que reproducen desigualdades preexistentes de clase, género y raza mientras las naturalizan bajo el lenguaje de la competencia individual. La datificación distribuye sus cargas de manera profundamente asimétrica, y los sujetos en condiciones de mayor vulnerabilidad sufren una exposición al control que los sujetos privilegiados pueden, al menos parcialmente, esquivar.

Zuboff (2019) elabora lo que la perspectiva haniana deja pendiente al mostrar que ese sujeto de rendimiento es simultáneamente explotado por terceros que extraen sus

datos de comportamiento como materia prima gratuita. Existe una asimetría radical entre quien produce los datos y quien los procesa, los vende y los convierte en predicciones comercializables, y la distancia entre la experiencia subjetiva de libertad y la realidad objetiva de extracción es, tal vez, la marca más definitoria del capitalismo de datos. La psicopolítica haniana, que ilumina con brillantez la dimensión subjetiva de este proceso, necesita articularse con una economía política de la extracción para dar cuenta de las condiciones estructurales que lo hacen posible a escala global.

## **Datos, plataformas y modelación algorítmica**

### *Smart power y conducción flexible de la conducta*

El concepto de *smart power* adquiere en el contexto de la tecnopolítica un significado preciso y diferente al de su uso original en los estudios de relaciones internacionales. En la era del control algorítmico, designa la capacidad de las arquitecturas digitales para conducir la conducta sin recurrir a la prohibición ni a la persuasión explícita, pues el poder inteligente actúa a través del diseño de entornos de acción, configurando los menús de opciones disponibles, ordenando los flujos de información, amplificando ciertas posibilidades y haciendo invisibles otras, todo ello de manera personalizada y adaptativa. Operar sobre probabilidades en lugar de sobre prohibiciones constituye su rasgo definitorio, y esa operación

marca una distancia cualitativa respecto de todas las formas previas de gobierno de la conducta.

Según Han (2014a), esta forma de poder actúa desde adentro del sujeto, haciendo propios sus deseos y orientándolos sin que el sujeto perciba la orientación, lo que explica que no genere resistencia porque se experimenta como libertad. El algoritmo de recomendación que selecciona el contenido que verá un usuario no le impide ver otros contenidos, pero hace que los seleccionados sean irresistiblemente más atractivos, relevantes y emocionalmente resonantes, de modo que el sujeto elige, en efecto, pero dentro de un campo de opciones que ya ha sido configurado para que ciertas elecciones sean estadísticamente más probables que otras. Thaler y Sunstein (2008) describen este mismo principio bajo el concepto de *nudge*, por el que la arquitectura de la elección se configura de modo que ciertas opciones se vuelvan más probables sin que el sujeto experimente ninguna restricción formal de su libertad. Lo que el *nudge* describe en el ámbito de las políticas públicas, el algoritmo de recomendación lo realiza a una escala incomparablemente mayor y con una precisión individualizante que el diseño institucional clásico nunca pudo alcanzar.

Lo que distingue al *smart power* de las formas anteriores de persuasión política es su carácter personalizado, predictivo y preemptivo. La propaganda clásica emitía un

mismo mensaje a toda la masa, y la publicidad moderna segmentaba la audiencia por grandes categorías demográficas, pero el *smart power* algorítmico construye un perfil individualizado de cada sujeto, identifica sus vulnerabilidades afectivas específicas y le presenta estímulos ajustados a su historia particular de interacciones digitales. El poder se vuelve radicalmente íntimo y actúa en el nivel de los miedos privados y de los deseos más sutiles de cada persona, concediéndole una eficacia que ninguna forma anterior de persuasión masiva podía alcanzar. La intimización del poder es, al mismo tiempo, lo que hace tan difícil articularlo como dominación, porque ocurre exactamente en el lugar donde el sujeto siente que es más libre. El poder más eficaz es aquel que se vuelve indistinguible de la voluntad propia del dominado.

Esta lógica articula el *smart power* con lo que Deleuze (1990) llamó modulación, pues a diferencia de la disciplina, que moldea al sujeto en un encierro institucional definido, el control modula continuamente al sujeto, ajustando en tiempo real los parámetros de su entorno de acción. La modulación algorítmica opera sobre perfiles estadísticos, anticipa la conducta futura y actúa sobre las posibilidades de la acción antes de que el sujeto delibere, de manera que el control puede ejercerse antes de la acción, como configuración anticipatoria del campo de lo posible, y no solamente después de ella como el castigo disciplinario. Esta temporalidad preventiva es decisiva porque el poder ya no necesita esperar



a que el sujeto actúe para corregirlo, le basta con configurar el espacio en que actuará para que la corrección resulte innecesaria.

La infraestructura material de este *smart power* son las plataformas digitales. Van Dijck (2013) ha mostrado que las plataformas transforman la sociabilidad en dato, y la distinción que propone entre usuario y perfil de usuario resulta filosóficamente decisiva, pues lo que la plataforma conoce del usuario no es a la persona con historia, afectos y proyectos, es a su perfil estadístico, y es con ese perfil con el que opera, decide y predice, sustituyendo progresivamente al sujeto por su representación algorítmica.

Crary (2013) observa que las lógicas del capitalismo de plataformas colonizan el tiempo de vida de los sujetos, haciendo que incluso el descanso y la inactividad se conviertan en objeto de valorización o en falla que debe ser corregida. La disponibilidad permanente del 24/7 es la forma temporal del *smart power*, pues prescribe una presencia continua en entornos que nunca cierran, que nunca permiten el silencio ni la interrupción, y esa prescripción ha sido internalizada como forma de vida por el sujeto contemporáneo hasta el punto de que la desconexión comienza a percibirse como déficit y no como derecho. El tiempo muerto, el tiempo no productivo, el tiempo que no genera datos se convierte en una anomalía que el sistema busca capturar o corregir, de modo que la colonización del

tiempo es la otra cara de la colonización de la subjetividad, porque un sujeto permanentemente disponible es un sujeto permanentemente datificable.

### *Vigilancia, predicción y economía de la conducta*

Zuboff (2019) elabora la arquitectura económica de este poder en términos de capitalismo de vigilancia. Las grandes corporaciones tecnológicas han descubierto que el verdadero recurso valioso no son los servicios que ofrecen a los usuarios, que en este sistema funcionan como cebo para la extracción, pues el recurso real son los datos de comportamiento que los usuarios producen al utilizarlos. Esos datos funcionan como materias primas que, procesadas mediante aprendizaje automático, permiten producir predicciones sobre la conducta futura, los llamados *behavioral futures*, que se comercializan en mercados donde anunciantes, aseguradoras, instituciones financieras y actores políticos pagan por el acceso a sujetos con determinada probabilidad de adoptar ciertos comportamientos. El sujeto es, en la lógica de este sistema, la materia prima de la que se extrae valor, aunque se lo trate ceremonialmente como cliente, pues la relación real no es de servicio, es de extracción. Sadowski (2019) profundiza este punto al mostrar que los datos operan como una forma de capital que se acumula, circula y produce poder, y no meramente como una mercancía que se compra y vende, configurando un régimen político-económico que él llama capitalismo de datos, en el que la

acumulación de información sobre la conducta humana se convierte en el vector principal de ventaja competitiva y control social.

La gubernamentalidad algorítmica rompe el modelo político moderno de manera fundamental porque opera sobre correlaciones estadísticas que los sujetos desconocen, mediante mecanismos que no fueron diseñados para ser comprensibles, y produce efectos sobre la vida de las personas sin pasar por ningún acto explícito de comunicación normativa. El gobierno ya no necesita hablar, justificarse ni persuadir, le basta con calcular, y ese cálculo silencioso constituye una forma de poder radicalmente nueva en la historia política moderna.

Rouvroy y Berns (2013) aportan la conceptualización más precisa de este mecanismo. La gubernamentalidad algorítmica opera sobre perfiles estadísticos contruidos a partir de datos infraindividuales, prescindiendo del sujeto como interlocutor y sin producir subjetivación. El sujeto de la gubernamentalidad foucaultiana debía conocerse a sí mismo, confesarse y reformarse, era un sujeto convocado a un trabajo reflexivo sobre su propia conducta. El sujeto de la gubernamentalidad algorítmica es, en cambio, un perfil que nunca se reconoce en las categorías que lo clasifican porque esas categorías son correlaciones estadísticas sin correlato fenomenológico en su experiencia vivida. El sujeto no sabe

que es una categoría para el sistema, y el sistema no necesita que lo sepa.

Couldry y Mejias (2019) añaden la dimensión del colonialismo de datos. La extracción masiva de datos constituye una relación de poder que apropia la experiencia humana como recurso explotable de manera estructuralmente análoga a como el colonialismo histórico se apropió de territorios y recursos naturales. El capitalismo de datos concentra el poder de conocimiento y predicción en pocas corporaciones radicadas en geografías específicas, privando a comunidades enteras de la capacidad de nombrar y resistir las formas en que su vida social es capturada y valorizada por terceros. Vista desde esta perspectiva, la datificación es también una reconfiguración de la geopolítica del poder epistémico.

Rouvroy y Berns (2013) subrayan que la posibilidad de emancipación frente a la gubernamentalidad algorítmica no puede residir en una reivindicación ingenua de mayor transparencia de datos. La emancipación pasa por reconstruir formas de subjetivación que el sistema algorítmico tiende a disolver, entre ellas las formas de singularidad irreductible al perfil, de imprevisibilidad irreductible a la predicción y de acción política irreductible al comportamiento estadísticamente esperado. La resistencia, en este marco, es la afirmación de lo que escapa al cálculo.

## **Ser, opacidad y crisis de la interioridad**

### *Del dato como registro al dato como forma de mundo*

El argumento ha procedido hasta aquí principalmente en un registro político-económico, analizando las condiciones estructurales y las lógicas de poder que producen al sujeto datificable. Conviene ahora descender al nivel ontológico, donde la cuestión adquiere su máxima densidad y su formulación más radical. ¿Qué significa ser cuando la existencia se ha vuelto sistemáticamente traducible en datos? Esta pregunta no equivale a preguntar por los efectos de la tecnología sobre el ser humano, como si el ser humano preexistiera intacto a la tecnología y esta simplemente lo afectara desde fuera, pues la cuestión es más radical. El proceso de datificación reconfigura la forma en que el sujeto se comprende a sí mismo, comprende a los demás y comprende el mundo en que habita, de modo que los datos comienzan a funcionar como mediación constitutiva de la experiencia, como la lente a través de la cual el mundo adquiere forma y sentido. El sujeto no usa datos de manera instrumental para luego volver a habitar un mundo previo, pues empieza a habitar un mundo cuya textura perceptiva y cuyas posibilidades de acción están cada vez más configuradas por la infraestructura de datificación que lo envuelve.

Esta afirmación requiere precisión para no volverse abstracta. El dato, en su sentido técnico, es una formalización de trazas producidas por la interacción entre sujetos y dispositivos digitales. Un *like*, un tiempo de permanencia en una página, una ruta de navegación, una pausa en la reproducción de un video. Ninguno de estos eventos es, en la experiencia del sujeto, un enunciado deliberado sobre sí mismo. Son gestos, hábitos, distracciones, curiosidades pasajeras, movimientos prerreflexivos que el sujeto realiza sin pensar en su significado y, muchas veces, sin recordarlos después. El sistema de extracción de datos los convierte, sin embargo, en elementos de un perfil que el sujeto no controla ni conoce, pero que produce efectos reales sobre sus posibilidades de vida. El dato no captura lo que el sujeto es, produce una versión de él que opera en registros de los que el sujeto está completamente excluido, una versión que toma decisiones en su nombre sin que exista ningún mecanismo por el que el sujeto pueda contestar, corregir o impugnar la representación que lo sustituye.

La economía de datos reorganiza también las capacidades cognitivas del sujeto de una manera que Stiegler (2020) describe como proletarianización cognitiva. Así como la revolución industrial despojó al artesano de sus saberes prácticos al transferirlos a la máquina, la automatización digital priva progresivamente al sujeto de los saberes teóricos y prácticos mediante los cuales se constituye como ser capaz de pensar y actuar de forma autónoma. Cuando los algoritmos

de recomendación seleccionan las rutas de información y los sistemas de predicción anticipan las decisiones del usuario, el sujeto pierde gradualmente la capacidad de orientarse por sí mismo en el espacio del conocimiento. Deja de saber buscar porque el sistema busca por él, deja de saber elegir porque el sistema elige por él, y ese despojamiento de las capacidades cognitivas fundamentales convierte al sujeto en un consumidor cada vez más dependiente de los sistemas que lo asisten. La captura industrial de la atención que Stiegler analiza no es un efecto colateral del capitalismo de datos, es su condición de eficacia, ya que un sujeto cognitivamente dependiente consume más, comparte más y resiste menos.

Han (2013) observa que la sociedad de la transparencia impone una lógica de visibilidad total que disuelve la opacidad necesaria para la singularidad. Lo que no puede ser cuantificado, perfilado y comparado tiende a ser invisible para el sistema, de modo que la opacidad es la condición de posibilidad de aquello en el sujeto que no puede ser reducido a perfil o promedio estadístico. El secreto, el silencio, la zona de sombra en que el pensamiento se forma antes de articularse son aspectos de una interioridad que, aunque no es anterior a lo social, tampoco es reductible a lo que ya existe como dato. Cuando esa zona de sombra es colonizada por la extracción de datos, se pierde el espacio en que el sujeto puede formarse, dudar y revisar sin que esos movimientos queden registrados y utilizados. La datificación erosiona ese espacio

gradual y silenciosamente hasta que el sujeto ya no advierte que lo ha perdido.

Rouvroy y Berns (2013) precisan este punto de manera filosóficamente rigurosa. La gubernamentalidad algorítmica opera sobre datos infraindividuales que son previos a cualquier acto reflexivo del sujeto, y a partir de ellos los sistemas algorítmicos construyen perfiles supraindividuales que predicen la conducta de conjuntos de sujetos similares, prescindiendo completamente del sujeto reflexivo como instancia relevante. El dato reemplaza al sujeto como objeto de gobierno, y esta sustitución tiene una consecuencia ontológica mayor. La realidad que importa para el sistema algorítmico es la realidad digital, la única que el sistema puede procesar, y esa realidad produce efectos materiales sobre la vida de los sujetos en su acceso al crédito, al trabajo, al seguro y a la información, sin que esos efectos pasen por ningún acto de comunicación, justificación ni reconocimiento mutuo. El sistema actúa sobre el perfil, la persona nunca es interpelada, y el perfil nunca puede responder, porque no es un sujeto, es una función estadística que opera en nombre de alguien que ignora ser representado por ella.

### *Libertad, juicio y autonomía bajo control algorítmico*

El efecto de esta configuración sobre la capacidad de juicio político es uno de los problemas menos explorados y más urgentes de la tecnopolítica contemporánea. El juicio, en



la tradición que va de Aristóteles a Arendt, es la capacidad de pensar en situaciones de indeterminación, de reconocer lo singular de la situación y de deliberar sin garantías previas. Es el ejercicio de pensar donde no hay regla previa que aplicar, y exige tiempo, distancia, memoria y apertura a la posibilidad de equivocarse, condiciones que el entorno algorítmico tiende a erosionar de manera sistemática.

Crary (2013) muestra que el capitalismo de plataformas produce una temporalidad de la inmediatez incompatible con las formas de atención sostenida que requieren la deliberación y la formación del pensamiento propio. El sujeto conectado permanentemente es un sujeto cuya atención es capturada de manera discontinua y fragmentada, cuya capacidad para tolerar la lentitud del pensamiento crítico se deteriora progresivamente. Ese deterioro no es accidental, pues un sujeto atento y crítico es, para los fines del capitalismo de vigilancia, menos rentable que un sujeto reactivo e impulsivo, revelando una tensión estructural que la retórica del empoderamiento digital sistemáticamente oculta.

La erosión del juicio tiene una dimensión específicamente política que merece atención particular. Cuando el entorno informativo es gestionado algorítmicamente de modo que maximiza la confirmación de sesgos previos y minimiza la exposición a perspectivas divergentes, el juicio político se ve privado de sus condiciones

de ejercicio, principalmente la pluralidad de voces, la confrontación con la alteridad y el tiempo para procesar la disonancia. El sujeto que solo recibe información confirmatoria pierde progresivamente la capacidad de imaginar que las cosas podrían ser de otro modo, y esa pérdida de la imaginación política es, quizás, la consecuencia más grave de la personalización algorítmica del entorno informativo.

Conviene, con todo, evitar el determinismo fácil. La datificación complica la posibilidad del juicio y la desplaza, pero no la cancela de manera absoluta, pues existen formas de apropiación crítica de las plataformas, prácticas de gestión consciente de la atención, comunidades de resistencia que utilizan las mismas infraestructuras digitales para organizarse y deliberar. La cuestión central es que las condiciones estructurales en las que se ejerce la agencia han sido significativamente modificadas por las arquitecturas algorítmicas en beneficio de los poderes que las administran, de modo que la libertad se vuelve más costosa, más difícil de ejercer y más fácilmente capturable, y esa modificación de las condiciones en que se ejerce la libertad es, en sí misma, una forma de poder sobre la libertad.

Han (2014a) señala, con razón, que el sujeto de la psicopolítica neoliberal experimenta su sujeción como libertad, y eso es precisamente lo que la hace tan eficaz y tan difícil de articular como opresión, dado que no hay un amo

visible al que oponerse, no hay prohibición que transgredir, no hay norma externa que rechazar, porque el poder opera a través de la libertad, la habita desde adentro, la configura como espacio de elección ya predeterminado. Este señalamiento conduce a una pregunta que Han tiende a eludir y que resulta decisiva. ¿Existe diferencia perceptible entre la libertad genuina y la libertad administrada algorítmicamente, y cómo puede el sujeto distinguir, desde adentro de la experiencia, si sus deseos son suyos o si han sido inducidos por un sistema de recomendación que conoce sus vulnerabilidades mejor que él mismo? Esta pregunta no tiene una respuesta individual. La respuesta es político-institucional y pasa por la regulación, la transparencia algorítmica y la construcción de contrapoderes capaces de hacer auditable el funcionamiento de los sistemas de predicción conductual, dado que la autonomía individual es una condición necesaria que no puede sostenerse sin las instituciones que la hagan posible.

La autonomía es, por tanto, una condición social que requiere ciertas estructuras institucionales, cierta distribución del poder epistémico y cierto equilibrio entre los actores que producen y los actores que controlan los sistemas de conocimiento sobre la conducta humana. Cuando ese equilibrio se rompe radicalmente, como ocurre en el capitalismo de datos donde la asimetría de información entre las plataformas y los usuarios es prácticamente absoluta, la autonomía se convierte en un ideal formal desprovisto de

condiciones reales de ejercicio. Proclamar la libertad del usuario en un contexto donde el usuario carece de los medios para comprender, evaluar o contestar los sistemas que lo gobiernan es una operación ideológica que enmascara relaciones de dominación bajo el lenguaje de la elección.

## **Consecuencias político-democráticas**

El problema del sujeto datificable es también, y tal vez ante todo, un problema político de primer orden, porque las transformaciones analizadas en los apartados anteriores afectan directamente las condiciones de posibilidad de la vida democrática. La democracia presupone ciudadanos capaces de formarse opiniones propias, de deliberar con otros sobre asuntos comunes y de actuar colectivamente sobre la base de juicios compartidos, y cada una de estas capacidades está siendo reconfigurada por las arquitecturas del control algorítmico de una manera que expresa intereses, lógicas de acumulación y relaciones de poder que no son neutras.

Brown (2015) ha mostrado que la racionalidad neoliberal transforma al ciudadano en capital humano, un sujeto que evalúa todas sus acciones en términos de rendimiento y competitividad, perdiendo la gramática política que permite pensar en términos de bien común o de justicia distributiva. El ciudadano que se piensa como capital humano no desaparece de la política, pero la política que practica se convierte en gestión estratégica de intereses

individuales, una extensión de la lógica del mercado al ámbito de lo público. A esta transformación subjetiva, el capitalismo de datos añade una dimensión técnica que la profundiza considerablemente. El sujeto vive en un entorno de información curada algorítmicamente que refuerza sus sesgos previos, limita la exposición a perspectivas divergentes y obstaculiza el encuentro con la alteridad que toda deliberación genuina requiere. La cámara de eco no es un efecto espontáneo de las preferencias de los usuarios. Es el resultado de algoritmos diseñados para maximizar la relevancia percibida del contenido a partir de los sesgos que ya existen en el comportamiento previo, de modo que el algoritmo no refleja al usuario, lo refuerza y, al reforzarlo, lo endurece.

Morozov (2013) ha denominado este proceso solucionismo tecnológico, con el que describe la tendencia a reencuadrar problemas políticos, éticos o sociales como problemas técnicos susceptibles de solución mediante herramientas digitales. Cuando los algoritmos se presentan como la respuesta a la gestión urbana, la salud pública o la participación política, los problemas genuinamente políticos quedan desplazados por preguntas de optimización que desactivan la dimensión normativa que les es constitutiva. Convertir un conflicto de valores en un problema de cálculo algorítmico equivale a suprimir el desacuerdo antes de resolverlo, vaciando el espacio político de su contenido propiamente normativo.

Como señalan Couldry y Mejias (2019), la datificación produce nuevas formas de desigualdad política invisibles en los términos clásicos del análisis democrático. La asimetría entre quienes pueden predecir y modular la conducta de millones y quienes ni siquiera saben que están siendo predichos es una asimetría de poder político que otorga ventaja estructural a los primeros sobre cualquier actor que opere sin esas capacidades. Esta asimetría no es una disfunción corregible, pues es el principio de funcionamiento del sistema, la condición que hace posible la extracción de valor y la acumulación de poder predictivo.

Han (2017) señala que la masa digital pierde la posibilidad de constituirse como sujeto político con memoria e identidad colectiva. El enjambre reacciona pero no delibera, ocupa espacios digitales pero no los habita políticamente, produce ruido pero no produce voz en el sentido aristotélico del término, aquella voz capaz de articular lo justo y lo injusto. Esta pérdida es inseparable de las lógicas de perfilamiento que fragmentan el espacio público en burbujas de relevancia individual, haciendo cada vez más difícil la experiencia de un mundo común que es el presupuesto básico de toda política democrática.

Han (2012) describe cómo el sujeto agotado por la autoexplotación carece de la energía y la distancia crítica necesarias para participar en la vida política de manera reflexiva. La economía de la atención agota la atención

precisamente donde más falta hace, en la deliberación pública y en el seguimiento sostenido de los asuntos colectivos. El cansancio político no es pereza ciudadana, es el efecto previsible de un régimen que extrae del sujeto toda su energía para la productividad individual y no le deja resto para la acción colectiva.

Couldry (2020) ha insistido en la necesidad de recuperar la crítica social como práctica teórica y política en un momento en que la datificación amenaza con disolver las condiciones que la hacen posible. Los sujetos humanos son agentes capaces de tomar distancia de las condiciones en que actúan y de actuar sobre ellas de manera transformadora, y esa capacidad reflexiva es una condición de posibilidad de la crítica misma y de cualquier proyecto de transformación. Recuperar la crítica en la era de la datificación exige reconstruir las condiciones para el ejercicio de esa agencia tanto en el plano teórico como en el institucional.

Hay que reconocer también que el espacio digital no es uniformemente un espacio de captura. Es también un espacio de organización, de producción de contrapúblicos y de articulación de movimientos que en otro contexto histórico habrían carecido de medios para coordinarse y hacerse visibles. Los mismos dispositivos que permiten el control algorítmico han permitido formas inéditas de acción colectiva y de resistencia. El problema teórico y político urgente es cómo producir las condiciones para que ese potencial de

reapropiación se realice de manera sostenida sin quedar absorbido por las lógicas de monetización y perfilamiento que caracterizan al capitalismo de plataformas.

Una regulación democrática del capitalismo de datos requiere, como mínimo, transparencia algorítmica sobre los criterios con que los sistemas automatizados toman decisiones que afectan a las personas, soberanía sobre los datos personales producidos en el uso de plataformas, y mecanismos de *accountability* con auditoría independiente y rendición de cuentas ante instancias democráticas. Ninguna de estas condiciones resuelve por sí sola el problema de fondo que este capítulo ha intentado articular, pero todas son necesarias para que ese problema pueda ser planteado como asunto político y no relegado como fatalidad técnica.

La dimensión democrática del problema conduce, en definitiva, a una pregunta de filosofía política que las teorías disponibles no resuelven satisfactoriamente. ¿Qué tipo de ciudadanía es posible cuando las condiciones de formación del juicio político han sido colonizadas por arquitecturas diseñadas para producir comportamientos predecibles y rentables? Esta pregunta no admite respuesta técnica, exige una reflexión sobre los fundamentos de la vida política democrática articulada con el análisis empírico del capitalismo de datos. La convergencia entre filosofía política y economía política de los datos constituye una de las tareas teóricas más urgentes del presente.



## Conclusiones

Este capítulo ha intentado mostrar que la datificación del sujeto es mucho más que un fenómeno de gestión de la información o de vigilancia digital. Constituye una reconfiguración profunda de la relación entre ser, poder y mundo que define una de las mutaciones históricas centrales del presente. El sujeto datificable es la figura emergente de un régimen de poder que produce sus propios sujetos, sus propias verdades y sus propias formas de obediencia, con una eficacia que proviene de que esas formas de obediencia son vividas como formas de libertad. Lo que hace de este régimen algo cualitativamente nuevo es la circularidad entre producción del sujeto y conocimiento del sujeto, la capacidad del sistema para producir al sujeto que dice conocer y para conocer al sujeto que ha producido, cerrando un circuito que hace cada vez más difícil distinguir entre lo que el sujeto es y lo que el sistema ha hecho de él.

La genealogía desarrollada en el capítulo muestra que esta figura emerge de desplazamientos articulados en los que las lógicas del poder se superponen y se refuerzan. Del sujeto disciplinario foucaultiano se transita al *homo economicus* de la gubernamentalidad neoliberal, producido activamente por una racionalidad que hace de la competencia y la inversión en sí mismo las formas universales de la subjetividad, como Foucault (2007) documentó en su análisis del curso de 1978-1979. De ese sujeto neoliberal se transita al sujeto de control

deleuziano, modulado de manera continua por redes y plataformas, y de este al sujeto datificable propiamente dicho, cuya subjetividad entera se convierte en materia prima de un capitalismo que ha aprendido a explotar no solo el cuerpo o el tiempo, también la psique, el deseo y la atención. El análisis de Lazzarato (2014) sobre la esclavitud maquínica subraya que esta extracción penetra en los niveles infraindividuales de la conducta, haciendo funcionar al sujeto como componente de una máquina que lo precede y lo excede sin necesitar su comprensión ni su adhesión.

El concepto de *smart power* ha permitido nombrar la lógica de este poder con precisión. Es el poder que seduce, perfila, anticipa y configura entornos de acción de manera tan fina y personalizada que el sujeto no puede percibir la diferencia entre su voluntad y la orientación algorítmica de su voluntad, y a diferencia del poder disciplinario, que prohíbe y castiga, o del poder normalizador, que vigila y clasifica, el *smart power* opera por dentro, sin dejar huella de coacción. Esta indistinción entre la voluntad del sujeto y la configuración del entorno que la orienta constituye la marca característica del control en la era del capitalismo de plataformas, la marca de un poder que ha aprendido a ser invisible porque ha logrado instalarse en el lugar exacto desde el que el sujeto cree ser más libre.

El análisis de Han es indispensable para comprender la dimensión subjetiva de esta mutación, pues su teorización

de la psicopolítica y la transparencia ilumina aspectos del problema que ni el análisis estructuralista de Foucault ni el análisis económico de Zuboff (2019) pueden capturar por separado. Han tiende, con todo, a una generalización abstracta que oscurece diferencias internas al fenómeno y presenta al sujeto como completamente capturado por la lógica del rendimiento, sin suficiente atención a las fisuras y resistencias que persisten en la experiencia concreta. El sujeto datificable es un sujeto cuya agencia ha sido complicada y cuyas condiciones de posibilidad han sido reconfiguradas, pero no anulada. Afirmar esa complejidad equivale, más bien, a reconocer que el poder, por extenso que sea, nunca agota por completo las posibilidades del sujeto sobre el que opera.

La dimensión colectiva de este problema ha emergido con particular fuerza en el análisis de las consecuencias político-democráticas. El solucionismo tecnológico que denuncia Morozov (2013) vacía el espacio político de su contenido normativo al convertir conflictos de valor en problemas técnicos susceptibles de optimización. La proletarianización cognitiva que analiza Stiegler (2020) priva al sujeto de los saberes que le permitirían orientarse críticamente en el espacio del conocimiento. La masa digital que describe Han (2017) pierde la capacidad de constituirse como sujeto político deliberativo con memoria e identidad colectiva. Las nuevas desigualdades epistémicas que documentan Couldry y Mejias (2019) convierten la asimetría de información en asimetría de poder político, reproduciendo

a escala global las lógicas extractivas del colonialismo histórico. Estas líneas de análisis, desarrolladas desde tradiciones teóricas diversas, convergen en un mismo diagnóstico, a saber, que la datificación no es un fenómeno que afecta a la política desde fuera, reconfigura las condiciones mismas en que la política es posible.

Frente a este panorama, la propuesta de Couldry (2020) de recuperar la crítica social como práctica fundada en la agencia reflexiva humana ofrece un punto de partida que este capítulo hace suyo. La datificación no cancela la capacidad crítica del sujeto, la desafía y la complica, exigiendo condiciones más robustas para su ejercicio. Esas condiciones son de orden teórico, porque se requieren marcos conceptuales capaces de articular la dimensión subjetiva y la dimensión estructural del problema sin reducir una a la otra. Son también de orden institucional, porque se exigen regulaciones, mecanismos de auditoría y formas de rendición de cuentas ante instancias democráticas. Y son de orden político, porque se demandan actores colectivos capaces de actuar sobre las condiciones de producción de la subjetividad digital con la lucidez suficiente para no reproducir, en su propia organización, las lógicas de datificación que pretenden contestar.

La pregunta que este capítulo deja abierta es, en última instancia, política. ¿Qué formas de vida común, qué instituciones y qué prácticas de subjetivación son capaces de

resistir la captura algorítmica y de producir condiciones reales para el juicio, la deliberación y la autonomía? Esta pregunta exige la convergencia entre la filosofía política, la sociología de las plataformas y la economía política de los datos. Lo que la filosofía puede hacer, y lo que este capítulo ha intentado hacer, es mostrar por qué esta pregunta es de primer orden y por qué pensar el sujeto datificable es pensar, hoy, una nueva configuración histórica del poder. La figura del sujeto datificable interroga la experiencia humana en la relación entre singularidad y clasificación, entre memoria y predicción, entre opacidad y visibilidad, entre la capacidad de actuar políticamente y la tendencia de los sistemas de anticipación conductual a volver el futuro meramente probable. Pensar esa figura con rigor es pensar, al mismo tiempo, el tipo de mundo que estamos construyendo y el tipo de sujeto que ese mundo está produciendo.

## Referencias

- Brown, W. (2015). *Undoing the demos: Neoliberalism's stealth revolution*. Zone Books.
- Couldry, N. (2020). Recovering critique in an age of datafication. *New Media & Society*, 22(7), 1135–1151. <https://doi.org/10.1177/1461444820912536>
- Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data colonizes human life and appropriates it for capitalism*. Stanford University Press.

- Crary, J. (2013). *24/7: Late capitalism and the ends of sleep*. Verso.
- Deleuze, G. (1990). Posdata sobre las sociedades de control. En *Pourparlers* (pp. 240–247). Minuit.
- Foucault, M. (1976). *Vigilar y castigar: Nacimiento de la prisión*. Siglo XXI. (Obra original publicada en 1975)
- Foucault, M. (2007). *Nacimiento de la biopolítica: Curso en el Collège de France (1978–1979)*. Fondo de Cultura Económica. (Obra original publicada en 2004)
- Han, B.-C. (2012). *La sociedad del cansancio*. Herder.
- Han, B.-C. (2013). *La sociedad de la transparencia*. Herder.
- Han, B.-C. (2014a). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder*. Herder.
- Han, B.-C. (2014b). *En el enjambre*. Herder.
- Han, B.-C. (2017). *In the Swarm: Digital Prospects* (E. Butler, Trad.). MIT Press. (Obra original publicada en 2013)
- Lazzarato, M. (2014). *Signs and machines: Capitalism and the production of subjectivity* (J. D. Jordan, Trad.). Semiotext(e). (Obra original publicada en 2010)
- Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. PublicAffairs.
- Rouvroy, A., & Berns, T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation: Le disparate comme condition d'individuation par la relation. *Réseaux*, 31(177), 163–196. <https://doi.org/10.3917/res.177.0163>

- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1), 1–12. <https://doi.org/10.1177/2053951718820549>
- Stiegler, B. (2020). *The Nanjing lectures 2016–2019* (D. Ross, Trad.). Open Humanities Press.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Ullauri Betancourt, S. A., Rivas Plata, A. M., & Cáceres Silva, A. V. (2024). Lo político en la contemporaneidad: Aproximaciones desde la psicopolítica de Byung-Chul Han. *Ius Humani. Revista de Derecho*, 13(2), 370–410. <https://doi.org/10.31207/ih.v13i2.377>
- Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

# Humanismo cívico: un legado de la filosofía política clásica para el mundo contemporáneo

*Civic Humanism: A Legacy of Classical Political Philosophy for the Contemporary World*

**Carlos Andrés Gómez Rodas<sup>37</sup>, Eddy Elvira Bedoya Argüelles<sup>38</sup>**

## Resumen

El texto explica los fundamentos y concreciones del humanismo cívico, propuesta fundada en la filosofía política de Aristóteles, especialmente, en su concepción teleológica de la naturaleza humana; frente al individualismo liberal moderno, el humanismo cívico reivindica la comunidad

---

<sup>37</sup> Universidad Pontificia Bolivariana. <https://orcid.org/0000-0001-5370-1431>

<sup>38</sup> Universidad Pontificia Bolivariana. <https://orcid.org/0009-0001-9353-5957>



política como ámbito natural de realización personal y moral y destaca la centralidad de la virtud, la amistad cívica y la noción de bien común, así como el papel de la familia y la escuela en la formación del carácter. En este sentido, el humanismo cívico ofrece una alternativa ética y política que articula libertad, verdad y comunidad desde una ontología realista y finalista.

**Palabras clave:** Humanismo cívico; filosofía política; mundo contemporáneo; ética; pensamiento aristotélico.

### **Abstract**

The text explains the foundations and concreteness of civic humanism, a proposal based on Aristotle's political philosophy, especially in his teleological conception of human nature; as opposed to modern liberal individualism, civic humanism claims the political community as the natural sphere of personal and moral fulfillment and emphasizes the centrality of virtue, civic friendship and the notion of the common good, as well as the role of the family and the school in character formation. In this sense, civic humanism offers an ethical and political alternative that articulates freedom, truth and community from a realistic and finalist ontology.

**Keywords:** Civic humanism; political philosophy; contemporary world; ethics; Aristotelian thought.

## Introducción

El presente trabajo se propone examinar los fundamentos filosóficos del humanismo cívico a partir de una relectura sistemática del pensamiento de Aristóteles, en particular, de su antropología, su metafísica y su teoría política; frente a los modelos liberales dominantes en la modernidad, centrados en el individuo autónomo, los derechos subjetivos y una noción negativa de libertad, el humanismo cívico reivindica una concepción teleológica y comunitaria de la vida humana cuyo marco natural de realización es la polis; esta postura no equivale, simplemente, a una restauración nostálgica del pasado clásico, y, por el contrario, constituye una alternativa crítica y propositiva que responde a las limitaciones del racionalismo moderno a nivel ético, político y educativo.

A través del análisis de categorías fundamentales (physis, eudaimonía, virtud y amistad cívica), este ensayo muestra cómo la tradición aristotélica ofrece recursos conceptuales sólidos para repensar la educación, la ciudadanía y el bien común desde una perspectiva integral; se destaca, en particular, la centralidad de la virtud como eje estructurador del carácter personal y del orden político, así como la importancia de las comunidades primarias (familia y escuela) en la formación moral de los ciudadanos; por último, se subraya la necesidad de una recuperación filosófica de la noción de ley natural, entendida no como imposición externa

sino como expresión racional de la naturaleza humana en aras de articular libertad, verdad y comunidad en una visión armónica y humanizadora de la vida social. Después de presentar la fundamentación metafísica y antropológica del humanismo cívica, se desarrollarán sus concreciones éticas y políticas.

Conviene especificar, en este mismo punto, el alcance del análisis. La crítica que aquí se lleva a cabo no se dirige, de manera indistinta, a toda tradición liberal, ni quiere ignorar las aportaciones históricas que la tradición del constitucionalismo, de la teoría de los derechos fundamentales y de la limitación del poder político han hecho a la acción pública. El hilo del debate se centra, por el contrario, en ciertas formulaciones del liberalismo moderno o contemporáneo —especialmente, en variantes del liberalismo individualista o del liberalismo procedimentalista— que construyen la comunidad política como una mera asociación instrumental orientada hacia la garantía de derechos subjetivos y a la coordinación de intereses privados.

En concreto, aparecen problemáticas tres suposiciones que son propias de ese horizonte: (i) una antropología del ser humano como individuo autosuficiente cuya identidad moral se define independientemente de sus vínculos constitutivos; (ii) una concepción normativa que da preferencia a la neutralidad estatal en contrapunto a las concepciones del

bien; y (iii) una visión negativa de la libertad, que es entendida primordialmente como no interferencia. No se tratará, en cambio, de cuestionamientos técnicos sobre la ingeniería constitucional, ni se presentará una reconstrucción pormenorizada de las diferentes corrientes liberales (clásica, libertaria o igualitaria), sino que se tomarán en cuenta el fondo antropológico y el fondo ético que hacen posible tales ordenamientos.

Desde aquí, el humanismo cívico no se presenta como una negación total de la modernidad política, sino como una oferta de fundación alternativa que procura recuperar la densidad ontológica de la vida común, puesto que muestra que la cuestión fundamental no es solo cómo limitar el poder, sino para qué existe la comunidad política.

## **Fundamentación metafísica y antropológica del humanismo cívico**

Son varias las definiciones del concepto de *humanismo cívico* que pueden encontrarse en la bibliografía especializada. Una primera y esencial la ofrece Llano, al inicio de un libro fundamental para toda investigación a este respecto:

Entiendo por “humanismo cívico” la actitud que fomenta la responsabilidad de las personas y las comunidades ciudadanas en la orientación y el

desarrollo de la vida política. Postura que equivale a potenciar las virtudes sociales como referente radical de todo incremento cualitativo de la dinámica pública. (Llano, 2015, p. 16)

Así pues, el humanismo cívico es la visión teórica y práctica de la vida social que da valor y fomenta tres características que se reclaman y se refuerzan mutuamente. Primero y más esencialmente, la centralidad de los hombres y mujeres reales y concretos, “que toman conciencia de su condición de miembros activos y responsables de la sociedad, y procuran participar eficazmente en su configuración política” (Llano, 2015, p. 16). En segundo lugar, la comprensión de las comunidades humanas como espacios irremplazables y definitorios en el desarrollo de las personas, quienes superan así las conductas individualistas, para actuar como ciudadanos con derechos y deberes inapelables. Finalmente, el humanismo cívico reivindica la importancia de la esfera pública, de hecho, porque:

no la concibe como un magma omniabarcante, sino como ámbito de despliegue de las libertades sociales y como instancia de garantía para que la vida de las comunidades no sufra interferencias indebidas ni abusivas presiones de poderes ajenos a ellas. (Llano, 2015, p. 16)

Son esos los rasgos fundamentales del humanismo cívico, y, como el mismo Llano (1999a) ha señalado, las raíces del humanismo cívico se encuentran en la filosofía política de Aristóteles, principal antítesis de la concepción política moderna. A diferencia del mecanicismo de la conciencia moderna, el planteamiento de Aristóteles ejemplifica la comprensión organicista típica de la era premoderna. En ese sentido, es necesario descomponer la filosofía política del Estagirita con el fin de identificar y distinguir sus elementos metafísicos, antropológicos, lógicos y conceptuales, tanto para comprender mejor la concepción política moderna como para entender con profundidad su alternativa tradicional (Miller, 2022).

Aristóteles es un filósofo sistemático, es decir, no solo ha desarrollado una filosofía que abarca casi todos los ámbitos de la vida humana, sino que, al mismo tiempo, todos los componentes de su sistema filosófico están contruidos para complementarse entre sí. De hecho, resulta llamativo verificar que algunos conceptos atraviesan todo el pensamiento aristotélico y se difunden en todos o casi todos sus campos. El más significativo de estos conceptos es *naturaleza* (*physis*). Desde la metafísica y la física de Aristóteles hasta la ética y la política, la naturaleza ha dado una continuidad y una fuerte coherencia al pensamiento aristotélico. La afirmación fundamental y definitiva de la concepción política de Aristóteles es que “el hombre es por naturaleza un animal político” (*Política*, I, 2. 1253a 2-3)

(Aristóteles, 2008). Aunque las lecturas contemporáneas de la filosofía política de Aristóteles suelen hacer hincapié en la sociabilidad y en la dimensión política del hombre, el elemento más importante y el sello distintivo del pensamiento político aristotélico es la naturalidad (Kayapinar, 2022).

Precisar el significado de naturaleza en el pensamiento aristotélico resulta significativo a la hora de considerar su importancia en el plano antropológico, ético y político de su filosofía. Si bien, ya en el libro II de la *Física*, el Estagirita elabora su noción de naturaleza, será el libro V de *Metafísica*, redactado a manera de diccionario, el que permitirá acercarse a esta noción capital de la filosofía antigua (Berti, 2022).

Son varios los significados que Aristóteles da al término *naturaleza* y todos se relacionan entre sí, pero puede decirse que dos de ellos se destacan por incluir a los demás y ofrecer las notas más sobresalientes del concepto. Según la primera definición, la naturaleza es “aquello de donde procede en cada uno de los entes naturales el primer movimiento, que reside en ellos en cuanto tales” (*Metafísica*, V, 4, 1014b 18-20) (Aristóteles, 1994), o sea, se está hablando aquí del principio inmanente del movimiento de los entes naturales. En la segunda definición, la naturaleza equivale a la esencia de los entes naturales, es decir, a aquello que los hace ser lo que son. En el primer sentido, la noción de naturaleza señala que los entes naturales tienen en sí mismos, y no como un añadido externo, el principio de su movimiento,

de este modo, las circunstancias exteriores serían solo condiciones de posibilidad para el despliegue de dicho dinamismo que tiende a un telos, a una finalidad (Sancho, 2022). Por otro lado, la segunda acepción descrita se identifica con lo que, usualmente, se denomina esencia, y que indica el conjunto de propiedades de un ente en cuanto pertenece a una clase o especie de objetos. Así pues, no solo la naturaleza indica las propiedades que permiten a un ente ser lo que es, sino que lo hacen pertenecer a una especie. Si cambian estas propiedades esenciales, el ente deja de ser lo que era y de pertenecer a la especie de la que antes hacía parte. Teleología e inmanencia son las claves para comprender el concepto aristotélico de naturaleza, pues las propiedades esenciales del ente determinan el comienzo, el desarrollo y el fin de los procesos que tienden a la perfección de dicho ente. En otras palabras, la naturaleza se refiere a propiedades esenciales cuyo sentido se explica atendiendo a la finalidad y perfección del ente natural. Ahora bien, si se considera el alcance del carácter inmanente y teleológico de la naturaleza en la esfera del saber ético, puede afirmarse que el bien del hombre como objeto de la pregunta ética consiste en el pleno despliegue de todas las tendencias que alberga potencialmente en sí y que se actualizan paulatinamente a lo largo de su vida (Jaeger, 1946).

Esta claridad sobre el concepto de naturaleza en Aristóteles es imprescindible para comprender su idea de la política. El término *polis* usado por Aristóteles no debe



entenderse exclusivamente supeditado al lugar histórico que tuvo en la Grecia Antigua. Tampoco es acertado decir que el filósofo se refiere al Estado o la ciudad como hoy se los entiende. Si bien es evidente que las características de la *polis* descrita por Aristóteles tienen como referente el contexto histórico al que su pensamiento se encuentra ligado, no hay que olvidar que él ve en ella elementos naturales que traspasan cualquier época.

En efecto, el concepto aristotélico de naturaleza no se obtiene promediando las características de los hombres, pues no es descriptivo, sino normativo y, más específicamente, teleológico (...).

Considerando el concepto aristotélico de naturaleza según esta dimensión teleológico-moral, se puede decir que la tesis de que el hombre es “político” por naturaleza significa, básicamente, que sólo en la polis y no en las demás comunidades el hombre puede “acabar su generación”, es decir, alcanzar su fin, llegar al pleno despliegue de sus potencialidades específicas y, con ello, a la felicidad. (Ossandón, 2001, p. 58)

No puede pensarse, entonces, que la propuesta política aristotélica ha sido superada porque la *polis* griega haya desaparecido, pues lo inquietante de su planteamiento radica en el carácter natural que se atribuye en él a la *polis*: “Resulta manifiesto que la polis es una de las cosas que son por

naturaleza (*tôn physei he pólis estí*)” (*Política*, I, 2. 1253a 2-3) (Aristóteles, 2008), lo cual resulta interesante a la luz del concepto de naturaleza expuesto en la misma obra.

Una cierta idea de hombre inspira una idea de sociedad, que termina permeando la cosmovisión de las personas, sus actitudes vitales y sus relaciones interpersonales. Es por ello que salir de la crisis implica confrontar los fundamentos teóricos en los que está cimentado el mundo occidental desde el siglo XVII, es decir, los planteamientos fundamentales del liberalismo político. La naturaleza política del hombre revela la conexión directa entre el perfeccionamiento del carácter y la vida ciudadana. No se puede ser más persona sin la convivencia con los otros, y, necesariamente, las virtudes o vicios del carácter trascienden la esfera de lo privado o individual, afectando el ámbito público. Son estas las ideas nucleares del humanismo cívico que inspiran un proyecto educativo (García, 2016).

La educación de la personalidad no se puede hacer aisladamente, porque se encuentra íntimamente relacionada con cuestiones que implican el trato con otras personas y la integración en grupos de diverso tipo. Por eso la formación del carácter posee una decisiva dimensión comunitaria, mal que les pese a los neoliberales individualistas, tan reacios a admitir que la educación es el punto más débil de sus teorías. El racionalismo liberal —la ideología dominante en la actualidad— no tiene modo humano de conciliar la educación

personal con la formación cívica. Y así se está produciendo una notoria desintegración ética en los llamados países del primer mundo, con las repercusiones en otras latitudes que resultan tan notorias como en el caso del terrorismo (Llano, 2015).

El pensamiento europeo del siglo XVII —cuando la modernidad en general, y la conciencia política moderna en particular comenzaron a florecer— se distinguió por su ruptura radical con las tradiciones metafísicas, cosmológicas, morales y políticas del pasado. El núcleo intelectual de estas tradiciones rechazadas era básicamente aristotélico. Ningún pensador, ni siquiera Platón, ha tenido tanto impacto en la historia intelectual e institucional de la Antigüedad y de la Cristiandad. Hubo siglos enteros en los que el mundo civilizado vivió a la sombra de Aristóteles, y no solo el mundo europeo, sino también el otomano y el africano; no solo el cristiano, sino también el judío y el islámico. Fueron siglos en los que el conocimiento universal estuvo contenido en los escritos de un solo hombre (Lerner, 1943).

El pensamiento de Aristóteles cimentó una tradición intelectual y dominó la mente civilizada durante dos milenios. Por supuesto, no todas y cada una de las creencias, juicios o actitudes de esta tradición fueron formulados por el propio Aristóteles, pero tales innovaciones, principalmente en sus detalles, se desarrollaron en el espíritu de los principios generales de la metafísica, la ética y la política aristotélicas.

También hay que notar que la mayoría de las ideas defendidas por Aristóteles no eran desconocidas antes de Aristóteles. Sin embargo, fue él quien, por primera vez, formuló sistemáticamente esas ideas en un sistema filosófico único y coherente (Kayapinar, 2022).

En consecuencia, por la representatividad de Aristóteles en esta antigua tradición, es justo hablar de “aristotelismo”. Se puede argumentar con seguridad que la modernidad se construyó sobre el rechazo radical del aristotelismo. Es cierto que no hubo otra cosmovisión que se volviera hegemónica en tantos contextos como el aristotelismo. Cuando la modernidad asaltó el viejo mundo, sus más perspicaces exponentes entendieron que era el aristotelismo lo que tenía que ser echado por tierra (MacIntyre, 1987).

También es interesante señalar que, aunque la disolución del aristotelismo comenzó con la comprensión aristotélica de la naturaleza, particularmente bajo los ataques de los innegables y revolucionarios descubrimientos de la nueva física y astronomía, esta se expandió rápidamente a su metafísica, su ética y su política. El fracaso de la física de Aristóteles —y, por supuesto, de su cosmología— frente a los descubrimientos de la nueva ciencia es bastante comprensible, pero el rechazo del análisis aristotélico de la acción humana no puede explicarse del mismo modo. Al parecer, la razón principal de la denuncia de las ideas ético-

políticas de Aristóteles en el siglo XVII fue el catastrófico desmoronamiento de su cosmología. Esto muestra que existe una fuerte correlación entre su metafísica, su física, su cosmología, su ética y su política. En consecuencia, ningún relato sobre el surgimiento de la concepción moderna de la política estará completo si no tiene en cuenta esa estrecha relación entre cosmología y política (Brague, 2011).

Conviene, pues, acudir a la teoría política de Aristóteles para comprender mejor el imaginario político moderno que es dominante hasta hoy y ver una alternativa a este. Sin embargo, en lo que sigue, no se pretende un recuento preciso del pensamiento político aristotélico sino identificar los rasgos paradigmáticos que le dan su identidad. Así, en lugar de tratar fragmentos específicos de sus obras, la síntesis se enfocará en conceptos y presupuestos específicos que conforman el marco general de su ética y su política.

Como ya se indicó, *naturaleza* es el concepto clave en la famosa máxima de Aristóteles “el hombre es por naturaleza un animal político” (*Política*, I, 2. 1253a 2-3) (Aristóteles, 2008). En primer lugar, el término *naturaleza* implica la imagen característica de la concepción organicista, a saber, suponer un movimiento autoinducido dentro de cada ente. Así, la *naturaleza* es el fin y el principio interno del movimiento y el cambio hacia ese fin. Este movimiento no se detiene hasta que el potencial se actualiza por completo. *Naturaleza*, en la terminología aristotélica, se refiere también

a la esencia de las cosas, es decir, a lo que hace que un ente sea ese ente y no otro, es decir, a aquello que hace que algo sea una cosa determinada.

En su primer sentido, la noción de naturaleza indica que los entes naturales llevan en sí mismos el principio de su movimiento, esto es, que su desarrollo proviene de virtualidades contenidas dentro de sí mismos, respecto de los cuales las circunstancias exteriores constituyen sólo condiciones necesarias para su despliegue.

Ahora bien, esta procedencia interna del movimiento natural no es suficiente para caracterizar la naturaleza en este primer sentido, pues más importante aún es referirse a su carácter teleológico (...) Para Aristóteles, en efecto, todo movimiento natural no sólo está inmanentemente regulado, sino también teleológicamente dirigido, es decir, el término hacia el que se dirige desde sí mismo el ente natural constituye su fin (*télos*) y éste se identifica con la plenitud, perfección o bien de dicho ente. (Ossandón, 2001, p. 35)

Todo lo que existe tiene una naturaleza y, por tanto, todo lo que llega a ser se mueve hacia la realización o actualización de su naturaleza. En consecuencia, el ser humano, como cualquier otro ente, lucha por la realización de

su naturaleza. La realización de los fines paso a paso en un orden jerárquico es inherente y naturalmente buena. El bien supremo es algo más allá de lo cual no hay ningún fin que actualizar. Este bien supremo que se desea por sí mismo y no por otra cosa es lo que se llama *felicidad*. Así, pues, la felicidad del ser humano reside en la actualización última de su esencia. La esencia del ser humano es la racionalidad. En otras palabras, el ser humano es esencialmente un animal racional y racionalidad significa elegir la opción razonable y actuar de forma razonable en cada circunstancia. La opción razonable, según Aristóteles, no es otra cosa que el bien humano o la virtud, definida por el filósofo griego como un término medio entre dos vicios —uno por defecto y otro por exceso de lo necesario en las pasiones y acciones—, determinado por la razón y por aquello por lo que decidiría el hombre prudente (*Ética Nicomáquea*, II, 6, 1106b35) (Aristóteles, 1998). Sin embargo, es claro que la vida del ser humano no se reduce a un solo acto o momento; por consiguiente, la actualización de la naturaleza humana significa tener una vida completa vivida de acuerdo con la virtud, lo que equivale a desarrollar un carácter virtuoso o ser virtuoso. Así, la felicidad humana consiste en tener una vida virtuosa.

En el camino de desplegar su naturaleza y esforzarse por actualizar su esencia (tanto física como espiritualmente), los seres humanos establecen unidades sociopolíticas distintas y cuyo grado de autosuficiencia difiere entre sí: la

familia, la aldea y el Estado (*polis*). En este camino universal y natural, la formación de la familia es el primer paso, bueno en sí mismo, pero insuficiente para la actualización de la naturaleza humana. Luego viene la aldea, superior a la familia y con un mayor grado de autosuficiencia, pero tampoco esta cumple todos los requisitos para que el ser humano sea pleno. Así, la naturaleza del ser humano sigue produciendo una energía interna que tiende a la constitución del Estado, que, según Aristóteles, es autosuficiente para desarrollar un carácter virtuoso. Esto significa que ninguna unidad social o política más allá del Estado es natural porque no queda en la naturaleza humana potencial que se actualice y produzca energía para un movimiento ulterior. Si se asume esta explicación organicista y teleológica, quienes viven en una unidad sociopolítica más pequeña o menos desarrollada que un Estado deberían ser considerados menos humanos. Como señala el propio Aristóteles:

Es evidente que la ciudad es una de las cosas naturales, y que el hombre es por naturaleza un animal social, y que el insocial por naturaleza y no por azar es un ser inferior o un ser superior al hombre. Como aquel a quien Homero vitupera *sin tribu, sin ley, sin hogar*. (*Política*, I, 2, 1253a) (Aristóteles, 2008)

El Estado de Aristóteles es un ser natural y una unidad orgánica que existe en aras de la vida buena en común, idea con la cual el Estagirita refuta al sofista Licofrón de Atenas, a



quien podría considerarse un precursor de la filosofía política liberal:

La ciudad no es una comunidad de lugar para impedir injusticias recíprocas y con vistas al intercambio (...) sino que esta es una comunidad de casas y de familias para vivir bien, con el fin de una vida perfecta y autárquica. (*Política*, III, 9, 1280b 12) (Aristóteles, 2008)

A esto, Aristóteles añade, a manera de corolario, que “una ciudad es la comunidad de familias y aldeas para una vida perfecta y autosuficiente, y ésta es, según decimos, la vida feliz y buena” (*Política*, III, 9, 1280b 14) (Aristóteles, 2008). Así pues, el entorno sociopolítico natural de los seres humanos es la *polis*.

Como puede verse, estas concepciones metafísicas, antropológicas, éticas y políticas son la antítesis de una cosmovisión liberal, cimentada en una metafísica que opone lo uno y lo múltiple, en una antropología que niega la naturaleza social del hombre y el carácter natural de la *polis*, en una ética constructivista donde lo real no tiene lugar alguno en la fundamentación del bien y en una filosofía política que reduce la vida ciudadana a fines funcionales y utilitarios.

Respondiendo al reto del filósofo Alejandro Llano, gran representante y teórico del humanismo cívico, la presente tesis doctoral problematiza las bases metafísicas, antropológicas, éticas y políticas de la educación liberal, en vistas a fundamentar filosóficamente una propuesta de educación escolar que se cimiente en los planteamientos centrales del humanismo cívico.

### **Traducción contemporánea: instituciones, democracia y esfera pública mediada**

La actualidad del humanismo cívico y su pertinencia en la contemporaneidad dependen de su potencial para traducir categorías clásicas a la organización institucional compleja. La polis antigua resulta irrepetible, pero el esquema normativo de la propuesta se puede reformular a la luz de una democracia constitucional.

Así, la noción de bien común puede releerse como el conjunto de condiciones institucionales que ofrecen a los ciudadanos la posibilidad de desarrollar capacidades cívico-morales (sistemas educativos capaces de asegurar el desarrollo moral, asociaciones intermedias, prácticas deliberativas y marcos normativos que protejan bienes relacionales esenciales); la amistad cívica puede considerarse como confianza pública y disposición personal a cooperar, y, especialmente, a deliberar razonablemente en contextos de pluralismo; y la prudencia (phronesis) remite a la capacidad

de discernir, en situaciones contingentes y heterogéneas, aquello que favorece efectivamente la vida buena compartida o la deliberación democrática frente a la fusión de lo estratégico.

La expansión de la esfera pública mediada por la tecnología digital presenta nuevos retos, como la aceleración del juicio y la influencia de las emociones en el debate, el debilitamiento de mediaciones comunitarias, etc.; en este contexto, el humanismo cívico enfatiza la necesidad de reconstruir espacios formativos intermedios —familia, escuela y asociaciones civiles— que contrarresten la tendencia de las tecnoestructuras.

Entendida así, la propuesta del humanismo cívico no consiste en volver a formas de otro tiempo, sino en ofrecer criterios normativos desde los cuales desarrollar una crítica a la manera en la que se vive la democracia contemporánea y el modo en el que se pueden fortalecer los aspectos éticos de las democracias.

### **Concreciones éticas y políticas del humanismo cívico**

Aristóteles, en el inicio de su *Política*, establece que toda acción humana se orienta hacia algún bien. Esta afirmación fundamenta la idea de que la vida en comunidad no es una excepción a esta regla; desde esta premisa se puede sostener que el humanismo cívico tiene profundas raíces en el

pensamiento político aristotélico porque ambos tienen como un pilar esencial que la virtud ciudadana no es un simple adorno o suplemento ético sino el principio esencial que distingue a la polis de otras formas de asociación como las alianzas militares o los acuerdos comerciales.

Aristóteles afirma, en el libro III de *Política*, que una ciudad verdaderamente tal debe ocuparse de la virtud; en caso contrario no se diferenciaría de una mera alianza utilitaria, y, en esta línea, se establece una estrecha relación entre racionalidad práctica y justicia: La primera solo se hace efectiva en el contexto de la polis, y, sin ella, el ser humano se degrada hasta convertirse en un animal salvaje; así, la virtud no es una disposición aislada sino una cualidad desarrollada en comunidad cuyo telos es la realización plena de la naturaleza racional y social del ser humano.

El principal desafío contemporáneo que enfrenta el humanismo cívico radica en la imposibilidad de reproducir la polis antigua en el marco del Estado liberal moderno; no obstante, Llano (2015) sostiene que ciertos elementos esenciales de la polis pueden adaptarse y desarrollarse hoy sin caer en una nostalgia anacrónica ni en reconstrucciones reduccionistas del orden social; el humanismo cívico, en lugar de enfocarse en los ideales modernos de libertad o igualdad, propone —siguiendo a Aristóteles— como el fin último de la vida humana, la *eudaimonia*, es decir, un estado de plenitud

y prosperidad que solo puede alcanzarse mediante el cultivo de las virtudes.

El humanismo cívico, contrario a la concepción kantiana que opone la moralidad a la inclinación, concibe la virtud como la armonía entre deseo y razón; en esta perspectiva la educación moral no es una represión de las pasiones sino una formación emocional profunda que orienta afectos y sentimientos hacia lo verdadero y lo bueno. Platón y Aristóteles coinciden en que la formación adecuada implica enseñar a amar lo que es digno de amor y a rechazar lo que es dañino, desde el comienzo de la vida hasta su madurez (Lewis, 2016).

La virtud como hábito racional no depende de la fortuna ni de los dioses; requiere, en cambio, juicio, discernimiento y capacidad de actuar correctamente en el momento y el lugar adecuados; por tanto no se reduce a la obediencia de normas externas; se cultiva en el marco de una comunidad donde cada individuo actúa en relación con los demás y con vistas al bien común; la comunidad, para lograr esta meta, debe promover la exaltación de las excelencias humanas y la inhibición de los vicios (MacIntyre, 1987).

El humanismo cívico, desde una perspectiva cognitivista, sostiene que los individuos pueden conocer lo que es bueno para su sociedad; Aristóteles, en esta dirección, define la virtud como término medio entre extremos viciosos,

una posición que exige actuar según la *recta razón* (*orthos logos*). Esta capacidad de juzgar es indispensable en la vida virtuosa y supera el mero cumplimiento legal (Llano, 2015; MacIntyre, 1987).

La prudencia (*phronesis*) ocupa un lugar privilegiado entre todas las virtudes; Aristóteles distingue entre virtudes intelectuales —adquiridas mediante la enseñanza— y virtudes morales —cultivadas a través del hábito—. Ambas, sin embargo, están profundamente entrelazadas; la unidad entre carácter y razón es condición para la vida ética, y, además, la virtud requiere un horizonte común, a saberse, el bien de todos los ciudadanos en tanto habitantes de la polis. Este bien común no es abstracto sino que se encarna en vínculos concretos como la amistad (Irizar, 2022).

En la concepción aristotélica, la amistad es una virtud, y, simultáneamente, el lazo que une a todas las demás; especialmente, significativa es la amistad personal virtuosa porque se basa en hábitos éticos compartidos y estimula el crecimiento moral mutuo; en contraposición, la sociedad liberal ha reducido la amistad a un sentimiento individual y contingente, perdiendo su carácter político; la Modernidad, al desarraigar a los individuos, ha favorecido una concepción de la vida pública como suma de intereses privados, olvidando la dimensión comunitaria y teleológica del ser humano (Carvalho, 1999).

La filosofía moral moderna levantó una barrera contra la ética de las virtudes bajo el influjo del individualismo moderno y su relativismo moral; en este marco se considera inviable una definición común de conducta moralmente recta; frente a este diagnóstico, el humanismo cívico propone superar esa barrera mediante un concepto central de virtud y un discurso moral comunitario; así, plantea que la conducta debe orientarse por valores compartidos y no por inclinaciones individuales o subjetivas. Superar el relativismo y concepciones afines (hedonismo, emotivismo, utilitarismo) se convierte en una tarea fundamental y la clave es redescubrir la posibilidad de una vida buena orientada hacia el bien común mediante un enfoque cooperativo (Llano, 2015).

El pensamiento de Hans Baron y su influencia en autores como Alejandro Llano permiten reformular el humanismo cívico en un contexto contemporáneo; Llano (2015) propone pequeñas comunidades orientadas al bien común como alternativa al individualismo liberal; Su *Humanismo cívico* representa una crítica al fracaso del liberalismo y del capitalismo para ofrecer un proyecto moral coherente. Frente al liberalismo (centrado en derechos individuales), al utilitarismo (centrado en el cálculo de beneficios) y a la deontología (centrada en el deber abstracto y desligado de la teleología), el humanismo cívico ofrece una vía intermedia, articulando libertad, virtud y comunidad (Irizar, 2007).

Una de las ideas centrales de esta propuesta es la redefinición de la libertad; el liberalismo la ha entendido como “libertad negativa”, es decir, ausencia de interferencia; esta noción —identificada por Isaiah Berlin (2001) y vinculada a Hobbes— desvincula la libertad de cualquier contenido moral o teleológico; sin embargo, para el humanismo cívico, la libertad no puede disociarse de la verdad ni de la naturaleza humana. Rechazar la noción de *physis* conduce a una ética sin fundamento ontológico, lo que equivale a sustituir la libertad por su caricatura postmoderna: una libertad sin dirección, meramente subjetiva y autodefinida.

En cambio, el humanismo cívico reintroduce la ley natural como expresión de la esencia del ser humano; esta ley no es impuesta desde fuera, sino que brota del mismo ser del hombre. La ley natural, lejos de oponerse a la libertad, le da dirección y sentido. La libertad humana es, entonces, la libertad de un ser creado, que encuentra su plenitud al actuar conforme a su naturaleza racional. Como han señalado autores como Pieper (2000) o Gómez Dávila (2005), sin la noción de creación, no es posible una antropología ética coherente (Gómez, 2016).

Desde esta óptica, el vínculo entre libertad y verdad es esencial; no se puede construir una teoría de la libertad sobre simples convenciones o percepciones culturales y una libertad sin anclaje ontológico es incapaz de sostener una vida ética



significativa. Con su “lógica del doble” y su rechazo de la verdad, el pensamiento postmoderno ha sustituido la libertad por la vacilación permanente entre imágenes vacías; frente a ello, el humanismo clásico enseña que la felicidad y la vida lograda solo son posibles mediante un uso recto y deliberado de la libertad, orientado hacia la plenitud personal (Llano, 1999b).

Tal plenitud solo se alcanza en el contexto de la comunidad política porque esta permite, no solo vivir, sino vivir bien; las leyes de la ciudad tienen como fin ordenar la vida hacia las virtudes. Aquí aparece la concordia o amistad cívica como otro componente fundamental que implica un acuerdo profundo entre ciudadanos sobre lo que es justo y beneficioso para todos; esta amistad política encuentra sus raíces en comunidades primarias como la familia y la escuela. Es en estas instituciones esenciales donde se aprenden los valores fundamentales de la ciudadanía, especialmente, la generosidad, el respeto y la benevolencia.

De ahí que el cultivo de una nueva sensibilidad ciudadana deba comenzar en el núcleo familiar; Llano propone cinco principios que deben guiar esta renovación: gradualidad, pluralismo, complementariedad, integralidad y solidaridad. La gradualidad enseña que la realidad es compleja y debe abordarse con espíritu crítico y diálogo; el pluralismo y la complementariedad recuerdan que las diferencias no deben generar conflicto, sino cooperación (la

mujer, en este sentido, tiene una especial capacidad para fomentar una cultura del encuentro y de la reconciliación); y la solidaridad es la clave de una convivencia auténtica; no puede ser impuesta como ley sino que debe surgir de una conciencia moral interior; esta virtud permite que las relaciones sociales se mantengan vivas y humanas, incluso, en medio de escenarios impersonales propios de la tecnoestructura. Las virtudes de los ciudadanos comunes son las que sostienen el edificio social; este fenómeno —que algunos llaman “revolución silenciosa”— marca un retorno a los vínculos solidarios y al protagonismo ético de las pequeñas comunidades (Llano, 1988).

### **Objeciones actuales y respuesta del humanismo cívico**

Una propuesta que reclama la centralidad de la virtud y del bien común en el espacio público, ciertamente, recibe objeciones considerables en el debate contemporáneo. Habría que atender al menos a tres de ellas.

#### *El pluralismo razonable*

Esto es así porque, en sociedades complejas y plurales, conviven teorías de la vida buena a veces dispares e incluso incompatibles. Desde este punto de vista, cualquier intento de orientar la comunidad política hacia un ideal sustantivo podría interpretarse como desconocer precisamente dicha

pluralidad. El humanismo cívico no desatiende el pluralismo de hecho; lo que cuestiona, más bien, es la tesis en virtud de la cual el pluralismo exige neutralidad axiológica extrema. Es, por el contrario, capaz de afirmar que todo tipo de comunidad política implica o presupone, en cuanto a los bienes compartidos que la sostienen —confianza, justicia, veracidad, responsabilidad—, el hecho de que la convivencia se convierte en imposible mientras no sean esos bienes los que se fomentan y se institucionalizan mediante la comunidad política. Tales bienes no se constituyen como una imposición perfeccionista arbitraria, sino que son condiciones mínimas de inteligibilidad práctica (de hecho).

### *La neutralidad estatal*

En filosofía política, hay algunas corrientes según las cuales el Estado no tiene que promover concepciones sustantivas de bien, sin embargo, la propia neutralidad es también tener una toma de posición antropológica: presupone que el individuo puede definir sus fines de manera independiente en relación con marcos comunitarios densos. El humanismo cívico va a otorgar una distinción: el poder político no tiene que imponer programas existenciales particulares, pero sí puede —y debe— favorecer las condiciones culturales y de educación que hagan posible el ejercicio responsable de la libertad.

## *El riesgo de paternalismo o perfeccionismo*

Existe la preocupación de que una ética de la virtud produzca coerción moral. Para actuar en contra de esta objeción, la tradición aristotélica identifica una diferencia entre la formación y la coacción: la virtud no la imponen de manera externa como un deber heterónomo, sino que se adquiere en virtud de hábitos libres en contextos comunitarios cargados de significado. Desde este punto de vista, la ley desempeña un rol pedagógico limitado: no produce directamente la virtud, pero puede generar un entorno para la misma.

Estas objeciones muestran que el debate no se resuelve simple y llanamente por oposición, sino que requiere de una suficiente precisión en los niveles y las competencias de las partes confrontadas. El humanismo cívico no propone sustituir el orden constitucional por una comunidad moral homogénea, sino enriquecer la arquitectura democrática con una comprensión más robusta sobre la formación del carácter y el sentido del bien común.

## **Conclusión**

El humanismo cívico, considerado como una relectura actual de la tradición aristotélica, lo que persigue es rescatar la centralidad de la virtud, del bien común y de las comunidades formativas en un contexto de endurecimiento

de la fragmentación cultural y de hipertrofia del individualismo. Su fortaleza radica en ofrecer una antropología relacional y teleológica que consienta una articulación entre la libertad y la verdad, sin llegar a disolver la pluralidad social.

Sin embargo, esta propuesta tiene, también, limitaciones reales: la complejidad institucional de los Estados modernos, la profunda diversidad axiológica o la globalización económica son circunstancias que obstaculizan la construcción de consensos sustantivos amplios. El humanismo cívico no elimina esas tensiones, pero proporciona criterios para orientarlas.

Entre esos criterios pueden anotarse: (i) el reforzamiento de la educación cívica orientada a la formación del carácter; (ii) la promoción de instituciones intermedias que encarnen bienes relacionales; (iii) el cultivo de prácticas deliberativas que integren una racionalidad y un respeto recíprocos; y (iv) un entendimiento de la libertad vinculado a la responsabilidad moral.

La tesis básica puede ahora formularse así: ninguna arquitectura institucional por sí misma puede ser suficiente si no se apoya en las virtudes ciudadanas efectivamente cultivadas. Solo a través de aquellas comunidades que aspiran a un bien común y a la búsqueda de un objetivo vinculado a la excelencia moral será posible sostener una democracia que no

sea una mera técnica de gobernar, sino una democracia con aspiraciones de ser una vida compartida verdaderamente humana.

## Referencias

- Aristóteles. (1994). *Metafísica*. Gredos.
- Aristóteles. (1998). *Ética Nicomáquea*. Gredos.
- Aristóteles. (2008). *Política*. Gredos.
- Berlin, I. (2001). *Dos conceptos de libertad*. Alianza.
- Berti, E. (2022). ¿Un equívoco en la tradición aristotélica? Las interpretaciones de *Metafísica* α 1, 993b 23-31. *Tópicos. Revista de Filosofía*, (62), 11–29. <https://doi.org/10.21555/top.v62io.1636>
- Brague, R. (2011). *La ley de Dios. Historia filosófica de una alianza*. Encuentro.
- Carvalho, H. (1999). *Tradição e racionalidade na filosofia de Alasdair MacIntyre*. Unimarco.
- García Norato, J. O. (2016). ¿De qué humanismo estamos hablando? En J. González Camargo (Ed.), *Atajos hacia el humanismo cívico* (pp. 15–26). Universidad Sergio Arboleda.
- Gómez Dávila, N. (2005). *Escolios a um texto implícito I*. Villegas Editores.
- Gómez Rodas, C. A. (2016). *Nicolás Gómez Dávila y la creaturidad: Una respuesta a los problemas en la fundamentación autonomista de la dignidad y los derechos humanos*. uca.edu.ar.

- Irizar, L. (2007). Humanismo cívico: Sus nudos conceptuales (II). *Civilizar. Ciencias Sociales y Humanas*, (7), 211–224.
- Irizar, L. (2022). Humanismo cívico: Transformar la cultura a través de las comunidades de amistad. *Cuaderno Cervantino*, (1), 139–150.
- Jaeger, W. (1946). *Aristóteles, bases para la historia de su desarrollo intelectual*. Fondo de Cultura Económica.
- Kayapinar, M. A. (2022). Aristotle's political theory: Metaphysics and physics meet ethics and politics. *Dîvân*, (1), 1–15. <https://doi.org/953492>
- Lerner, M. (1943). *Introduction to Aristotle's Politics*. Random House.
- Lewis, C. S. (2016). *La abolición del hombre*. Harper Collins.
- Llano, A. (1988). *La nueva sensibilidad*. Espasa Calpe.
- Llano, A. (1999a). El humanismo cívico y sus raíces aristotélicas. *Anuario Filosófico*, (32), 443–468.
- Llano, A. (1999b). Tres formas de libertad. En J. Aranguren (Ed.), *La libertad sentimental* (pp. 11–30). Servicio de Publicaciones de la Universidad de Navarra.
- Llano, A. (2015). *Humanismo cívico*. Ediciones Cristiandad.
- MacIntyre, A. (1987). *Tras la virtud*. Crítica.
- Miller, F. (2022). A teoria política de Aristóteles. *Política*, (1), 167–207. <https://doi.org/10.23925/politica.v10i1.60016>

- Ossandón, J. C. (2001). *Felicidad y política. El fin último de la polis en la filosofía de Aristóteles*. Servicio de Publicaciones de la Universidad de Navarra.
- Pieper, J. (2000). Creaturidad y naturaleza humana. Notas sobre el planteamiento filosófico de Jean Paul Sartre. En *La fe ante el reto de la cultura contemporánea* (pp. 254–268). RIALP.
- Sancho Rocher, L. (2022). Aristóteles y la posibilidad de una mése politeía. *Araucaria. Revista Iberoamericana de Filosofía Política, Humanidades y Relaciones Internacionales*, (49), 351–372.  
<https://doi.org/10.12795.17>



# **Desigualdad neoliberal, neofascismo, tecnofeudalismo y necropolítica. Claves para disputar la democracia**

*Neoliberal Inequality, Neofascism, Techno-Feudalism, and Necropolitics. Keys to Challenging Democracy*

**Julio Ernesto Peña y Lillo Echeverría<sup>39</sup>**

*El neoliberalismo ha llevado a una concentración de riqueza que ha dejado a millones en la pobreza y ha erosionado el tejido social.*

Stiglitz (2002)

---

<sup>39</sup> Facultad Latinoamericana de Ciencias Sociales.  
<https://orcid.org/0000-0003-2201-2021>

*La necropolítica se hace visible en el neoliberalismo cuando el Estado abandona a sectores enteros de la población a la precariedad, al despojo o a la violencia.*

Mbembe (2011)

## **Resumen**

El trabajo que presentamos a continuación, aborda la problemática de la desigualdad neoliberal como una decisión política deliberada, que configura las relaciones de poder y distribución de recursos en las sociedades contemporáneas. Se examina cómo el neoliberalismo, lejos de ser un fenómeno económico neutral, produce exclusión estructural y vulnerabilidad social. A partir de esta base, se analizan el neofascismo, la necropolítica y el tecnofeudalismo, como sus mutaciones extremas, formas autoritarias de gestión social que profundizan la exclusión y fragmentan la democracia. Finalmente, el trabajo sostiene la necesidad urgente de llevar a cabo una recuperación del Estado, con la cual será posible contrarrestar las crisis generadas por el neoliberalismo, garantizar los derechos ciudadanos y construir sociedades más justas, inclusivas y democráticas.

**Palabras clave:** Neoliberalismo; neofascismo; necropolítica; tecnofeudalismo desigualdad; pobreza; Estado; derechos; inclusión; democracia.

## **Abstract**

The paper we present below addresses the problem of neoliberal inequality as a deliberate political decision that shapes power relations and resource distribution in contemporary societies. It examines how neoliberalism, far from being a neutral economic phenomenon, produces structural exclusion and social vulnerability. From this basis, it analyzes neofascism, necropolitics, and technofeudalism, as their extreme mutations—authoritarian forms of social management that deepen exclusion and fragment democracy. Finally, the paper argues for the urgent need to carry out a recovery of the State, which will make it possible to counteract the crises generated by neoliberalism, guarantee citizens' rights, and build more just, inclusive, and democratic societies.

**Keywords:** Neoliberalism; neofascism; necropolitics; technofeudalism; inequality; poverty; State; rights; inclusion; democracy.

## **Introducción**

Desde los años 80 el neoliberalismo ha tenido una gran influencia en las economías y sociedades del mundo contemporáneo, así como también en las estructuras políticas y en las formas de gobernanza, es decir, en las maneras en que se ejerce el poder y en que se toman las decisiones.

Lejos de ser una ideología netamente económica o técnica, el neoliberalismo se ha consolidado como una cultura político-económica, que pretende posicionar la idea de que el juego de las fuerzas económicas privadas, puede generar mejores resultados que cualquier intervención pública. Esto explica las políticas de reducción del Estado y con ello, de debilitamiento de los sistemas de protección social, servicios públicos (salud, educación, jubilaciones), la falta de inversión en infraestructuras, etc., medidas que han contribuido con la expansión y crecimiento de la desigualdad, la pobreza y la violencia social, a niveles sin precedentes.

Podríamos decir que se trata de una estrategia deliberada, que reconfigura las relaciones de poder, anteponiendo los intereses corporativos sobre el bien común, cooptando el Estado para el cumplimiento de sus fines, reduciendo o retirando al mismo tiempo, las capacidades estatales de protección y garantía de los derechos sociales fundamentales. En este sentido, el neoliberalismo en tanto que régimen político-económico, redefine la justicia social y el desarrollo ciudadano, a partir de los parámetros de mercado.

Sin embargo, su desenvolvimiento en el tiempo no ha permanecido estático, ha experimentado mutaciones que reflejan nuevas formas de control social y político. Entre estas transformaciones sobresalen fenómenos como el neofascismo, la necropolítica, el tecnofeudalismo, que pueden

entenderse como extensiones autoritarias y violentas utilizadas para garantizar su continuidad y permanencia.

Frente a este complejo escenario, surge una pregunta crucial: ¿cómo recuperar y revalorizar el rol del Estado, como garante efectivo de los derechos y de las condiciones necesarias para poder alcanzar una vida digna en democracia?

Las propuestas que presentamos en el ensayo teórico a continuación, son una invitación a repensar, a partir de los aportes y las contribuciones de la filosofía política, el papel que puede jugar el Estado, más allá de la caricatura de su función subsidiaria o reguladora, limitada por el mercado. El esfuerzo que se propone esta investigación, es el de recuperar el valor de sus funciones, como actor central en la redistribución –equitativa- de recursos, en la protección social integral y en la promoción activa de los derechos ciudadanos.

Para poder confrontar y reducir las brechas históricas de desigualdad, el racismo, la violencia y la exclusión, reproducidas y agravadas por el neoliberalismo y las formas autoritarias que lo acompañan, consideramos indispensable en nuestros tiempos, reexaminar el aporte de pensadores y teóricos políticos, que han contribuido a poner en valor el sentido de un Estado presente, activo y democrático, capaz de garantizar el desarrollo de una sociedad más justa e inclusiva.

El marco teórico conceptual que presentamos a continuación, constituye en nuestros días, un acto vivo de resistencia, que busca aportar a un debate urgente sobre los desafíos políticos contemporáneos y las posibilidades reales que tenemos como sociedad, como ciudadanos e individuos, de construir democracia, en tiempos marcados por profundas crisis económicas, sociales y humanitarias.

### **La desigualdad –neoliberal- es una decisión política**

En nuestros días el crecimiento de la desigualdad, la pobreza o la violencia, no puede comprenderse únicamente como una consecuencia de las dinámicas propias al mercado (ley de oferta y demanda), como pretenden hacernos creer los defensores del neoliberalismo. Por el contrario, como plantea Joseph Stiglitz (2012), la desigualdad y la reproducción de la pobreza son el resultado directo de decisiones institucionales y normativas tomadas a favor de los sectores más acaudalados.

El neoliberalismo lejos de ser una simple doctrina económica, es un proyecto político de reconfiguración cultural, que pretende normalizar o naturalizar un rol del Estado completamente sometido o subordinado a los intereses del capital. Se trata de una ideología hegemónica que promueve “más mercado y menos Estado” (Friedman, 1962; Hayek, 1944), con el fin de privatizar lo público y

desregular el mercado financiero, para acomodar la política a los intereses de las élites económicas.

El postulado ideológico neoliberal esconde, como nos recuerda David Harvey (2005), un proceso sistemático de "acumulación por desposesión", que implica la privatización de los bienes comunes, la mercantilización de los servicios públicos, el endeudamiento forzoso de los sectores populares, la expulsión de comunidades de sus tierras, bajo la apariencia de neutralidad técnica. De esta manera el neoliberalismo reconfigura al Estado, ya no como garante del bien común, sino como instrumento de reproducción del poder de clase.

Esta simbiosis entre capital y poder político desdibuja los límites entre lo público y lo privado, despojando a las grandes mayorías de sus derechos, servicios y recursos fundamentales para poder llevar una vida digna. El sector privado en el neoliberalismo absorbe las funciones del Estado democrático, sin necesidad de violencia, sin ruptura institucional, a través de la cooptación del Estado, sus activos y sus servicios públicos (Sheldon, 2010), generándose de esta manera un vaciamiento progresivo del contenido sustantivo de la democracia.

El crecimiento económico del sector privado o de las corporaciones no se traduce en beneficios para los ciudadanos, puesto que los sectores del capital se sirven de paraísos fiscales y reconfiguraciones legales para poder evadir

impuestos, despedir libremente a los trabajadores, reducir salarios o derechos adquiridos (como liquidaciones), perjudicando la economía de las familias, afectando gravemente los principios republicanos de solidaridad nacional, o la cultura tributaria progresiva del que más tiene más aporta.

Con el neoliberalismo el Estado deja de ser garante del bien común, de los derechos y de las condiciones de vida. Los servicios públicos se convierten en un botín que sirve de herramienta de acumulación al sector privado, provocado una redistribución de la riqueza a la inversa, desde las clases bajas y medias hacia las élites económicas, afectándose gravemente las estructuras de solidaridad, la protección social, los derechos colectivos, mientras se impone una visión individualista y competitiva que genera un mayor sufrimiento social (Harvey, 2005).

Este modelo no solo degrada las condiciones materiales de existencia, sino que debilita los principios fundantes de la vida democrática. Como afirma Thomas Piketty (2014), el neoliberalismo transforma la desigualdad económica en desigualdad política: quienes concentran la riqueza, acumulan también la capacidad de influir, presionar y moldear las decisiones públicas. Así, la democracia representativa se convierte en una farsa: el *demos* es sustituido por la plutocracia, y el ideal de igualdad de



oportunidades queda relegado frente a un orden estructural que antepone el privilegio.

La concentración del poder económico se traduce, inevitablemente, en un dominio político excluyente. Las élites financian campañas, imponen marcos normativos y bloquean cualquier intento de reforma redistributiva. El Estado deja de ser árbitro entre clases sociales, para convertirse en garante de la acumulación corporativa. En palabras de Bolívar Echeverría (2010), esto explica la configuración de sociedades a dos velocidades o una forma de apartheid que escinde el cuerpo político entre una minoría convocada y una mayoría desechada.

Este fenómeno adquiere dimensiones particularmente violentas en América Latina, donde el neoliberalismo perpetúa la exclusión histórica de pueblos indígenas, afrodescendientes y sectores populares, reproduciendo las lógicas coloniales de desposesión y marginalización.

Esta configuración política no es solo injusta, es inviable. El deterioro de la calidad de vida y la precarización del trabajo provocan una contracción de la demanda agregada, lo cual lleva al cierre de negocios, despidos masivos y un aumento de la criminalidad.

La racionalidad neoliberal, al despojar a los individuos de su dimensión ciudadana, los convierte en gestores aislados

de su propia supervivencia. Se impone una lógica meritocrática cruel, en la que el fracaso individual es responsabilizado al sujeto, ocultando las causas estructurales de la exclusión. El resultado es una sociedad fragmentada, donde la competencia sustituye a la solidaridad, y el miedo sustituye a la esperanza. Con el neoliberalismo se impone una tecnología de poder que socava las bases materiales y la posibilidad de un proyecto común, solidario y justo.

### *Neoliberalismo como ausencia de reconocimiento*

La desigualdad, a más de ser una consecuencia histórica y económica, en nuestros días es, ante todo, un problema político relacionado con la falta de reconocimiento, redistribución y participación. Sociedades injustas, nos dice Honneth, (1997), son aquellas que niegan, invisibilizan, excluyen y humillan a grandes sectores de la población.

Sin reconocimiento por parte del Estado, no hay políticas públicas, ni servicios públicos que permitan que los individuos se constituyan como sujetos íntegros, libres, y auto determinados, con posibilidad de llevar una vida digna (Honneth, 1997).

Una sociedad justa nos dice Honneth (1997), es aquella que permite que cada individuo pueda ser reconocido como igual, en dignidad y en derechos, o, en otras palabras, que nadie viva sin ser visto, sin ser escuchado, sin ser tratado

como sujeto valioso. Si las personas no reciben el reconocimiento, a través de políticas que les permitan pasar de un estado inhumano, a un estado de vida digna, transformando de esta manera una realidad injusta, se produce lo que Honneth llama, un “daño moral” o un “desprecio político”, que se traduce en una forma de violencia sistémica que afecta la salud psíquica y emocional de los individuos.

El neoliberalismo para Han (2024), no solo transforma las instituciones, sino que penetra en la subjetividad, genera angustia, estrés y miedo entre los ciudadanos, porque impone una lógica de competencia permanente, de inseguridad estructural y de autoexigencia constante. Factores que se ven agravados con el despojo del sentido colectivo, de las redes de protección social y de la estabilidad vital, lo cual se traduce en un sentimiento de abandono, que moldea cómo las personas viven, sienten y se relacionan consigo mismas y con los demás.

El neoliberalismo extrae derechos a los ciudadanos, procurando que cada persona se convierta en empresaria de sí misma, responsable de su éxito o fracaso, a pesar de las condiciones estructurales e históricas de desigualdad. La sociedad neoliberal promueve que los individuos se auto-exploten a sí mismos, en nombre de la libertad y de una supuesta realización personal (Han, 2024). Esto promueve

sociedades de individuos que se vuelven sus propios verdugos al incrementar los estados de fatiga, ansiedad y depresión.

La negación del reconocimiento, y con ello, la reducción o ausencia del Estado en el ámbito de desarrollo social, no es entonces un problema meramente individual o psicológico; por el contrario, se trata de una patología estructural de la sociedad. El Estado mínimo lejos de propiciar mayores libertades entre los ciudadanos, con su ausencia de políticas y de presencia los despoja, confiscando las posibilidades de superación o impidiendo su autorrealización.

La democracia, para Honneth (1997), debe garantizar condiciones institucionales y culturales de reconocimiento de todas y todos los ciudadanos. Las políticas públicas deben enfocarse en responder las necesidades de las mujeres, los pueblos indígenas, las personas racializadas, los trabajadores informales, las juventudes, las disidencias sexuales, las personas con discapacidad, etc. Una democracia sin reconocimiento, es una democracia vacía, incapaz de generar pertenencia, cohesión social o esperanza colectiva.

La falta sistemática y estructural de reconocimiento, conlleva a la invisibilización, la humillación, y la exclusión de grandes sectores de la sociedad, que, en ausencia de Estado, viven experiencias de falta de consideración, desprecio e irrespeto. El reconocimiento para Axel Honneth (1997), no

puede ser de uso exclusivo de los grupos corporativos, sino parte de una política pública encaminada a promover y garantizar una mayor paz social, salud psíquica individual y el desarrollo de sociedades más justas y equitativas.

Recuperar el reconocimiento como categoría política, no es solo una tarea teórica, sino un imperativo político, encaminado a reducir las brechas de la desigualdad, reconstruir los lazos rotos por el mercado y profundizar-expandir la democracia, como una práctica de inclusión radical, frente a la normalización política-neoliberal del desprecio.

### *Naturalización de la desigualdad o despolitización de la injusticia*

En las sociedades regidas por la lógica neoliberal, la desigualdad ya no es vivida como una afrenta moral, ni como un escándalo colectivo. Se ha convertido en una realidad “naturalizada”, percibida como parte inevitable del orden social.

Esta aceptación resignada de la injusticia, como advierte François Dubet (2021), no es un proceso espontáneo, sino el resultado de una profunda transformación cultural y política, cuyo objetivo es desactivar la capacidad de indignación, inhibir el conflicto y bloquear toda imaginación emancipadora.

El neoliberalismo, más que un modelo económico, es una forma de gobierno que produce subjetividades resignadas, fragmentadas y culpabilizadas. La desigualdad en este marco, deja de ser entendida como consecuencia de estructuras injustas o decisiones políticas deliberadas. En palabras de Dubet (2021), se borran las huellas de la clase social de origen, del género, de la raza, del territorio, del acceso desigual a los capitales económico, social o cultural (Bourdieu, 1986).

Al ocultarse estas estructuras de dominación, se desplaza la responsabilidad hacia los individuos, imponiéndoles la carga de su propio fracaso en un sistema diseñado para producir ganadores y perdedores de forma sistemática. Esta operación ideológica tiene un correlato evidente en los medios de comunicación, que, en contextos como el ecuatoriano, lejos de denunciar las condiciones que perpetúan la pobreza, la violencia o la exclusión, se convierten en cómplices de las condiciones de expoliación y exclusión trazadas por las políticas económicas de las élites (Bourdieu, 1986).

Televisión, radio, prensa y ciertas plataformas digitales operan como máquinas de normalización de las injusticias: no se enfocan en el sufrimiento social, desacreditan las demandas populares, criminalizan la protesta y promueven una visión meritocrática que glorifica a los exitosos (por herencias) y desprecia a los excluidos (por sistema).

Cuando las injusticias estructurales se presentan como “naturales”, se desactiva la capacidad colectiva de organización y resistencia. Las personas dejan de ver a los otros como iguales, para verlos y percibirlos como rivales o amenazas, minando directamente las bases de la solidaridad, la redistribución o el cuidado colectivo. En este escenario, la gente ya no lucha por derechos comunes, sino por mejorar su posición individual dentro de una jerarquía social que reproduce las injusticias sociales.

Esta erosión del vínculo social conlleva a graves consecuencias políticas que terminan minando los fundamentos republicanos y democráticos, al debilitar las nociones de comunidad, ciudadanía y derecho. La pérdida de solidaridad fragmenta a la sociedad, impide que se desarrollen las luchas colectivas, favoreciendo la cultura del individualismo competitivo, propio de la cultura neoliberal (Dubet, 2021).

En lugar de un “nosotros” capaz de construir proyectos colectivos, emerge un conjunto de individuos atomizados, rivales entre sí, desconectados de cualquier horizonte de igualdad o justicia. La lucha política se sustituye por el cálculo individual, y la solidaridad se transforma en sospecha. En este contexto, cualquier intento por restituir el principio de igualdad es percibido como una amenaza al orden naturalizado, y sus portadores: movimientos sociales, líderes

populares, gobiernos progresistas, son perseguidos, estigmatizados o judicializados.

La tragedia de este modelo radica en que convierte lo político en técnica, la justicia en mérito, y la vida común en competencia. En lugar de pensar la sociedad como destino compartido, el neoliberalismo impone una visión del mundo donde cada uno debe salvarse por su cuenta, renunciando a la idea de garantizar una vida digna para todos.

La desigualdad, la pobreza, la violencia sistémica en el neoliberalismo dejan de ser tratados como un problema común, para convertirse en un asunto privado, reducido a la capacidad de las trayectorias personales.

Esta perspectiva oculta el peso de las herencias, el origen social, el capital cultural, las estructuras económicas y políticas, profundizando y legitimando las injusticias sociales, reproduciendo sociedades del miedo, de la angustia, del estrés, la frustración, y de la culpa en quienes no logran mejorar su situación (Dubet, 2021; Han, 2024).

Repolitizar la desigualdad significa comprender que la pobreza, la violencia estructural y la exclusión no son accidentes, sino efectos de un orden que puede y debe ser transformado. Significa, en última instancia, volver a pensar la política como espacio de justicia, de solidaridad, y de construcción de un futuro común.



## **Neofascismo y las mutaciones autoritarias del neoliberalismo**

Vivimos una época marcada por una profunda mutación de las formas de poder, en donde el neoliberalismo se radicaliza bajo nuevas formas autoritarias, que conjugan el individualismo extremo, con el desprecio por lo común. Estamos presenciando, como advierte el filósofo brasileño Vladimir Safatle (2025b), el resurgimiento de una forma renovada de fascismo, que no se presenta con los emblemas del pasado, sino disfrazado de libertad individual, defensa de la propiedad privada y cruzadas morales contra los derechos sociales y colectivos. Esta nueva configuración del autoritarismo se despliega en figuras como Donald Trump, Javier Milei o Daniel Noboa.

El presidente Trump representa en nuestros días, una amenaza global para la democracia, al impulsar políticas autoritarias y expansionistas. En los últimos meses hemos podido apreciar la estigmatización, hostigamiento y persecución a los migrantes, a través del Servicio de Inmigración y Control de Aduanas (ICE), que ha convertido a estos grupos en foco de amenazas y víctimas del acoso policial (Primicias, 2026).

Entre el 1 y el 26 de enero de 2026, EEUU deportó a 742 ciudadanos ecuatorianos según cifras oficiales de la Cancillería del Ecuador. En el año 2025 fueron deportados un

total de 9.351 ecuatorianos desde Estados Unidos, en el marco de las políticas migratorias implementadas por el gobierno estadounidense, que incluyeron redadas del Servicio de Inmigración y Aduanas (ICE), deportaciones exprés y salidas “voluntarias” (Primicias, 2026). Muy similar a como operaban los nazis con la persecución a los judíos.

En materia de política internacional, el trumpismo se ha caracterizado por un constante desprecio al Derecho Internacional y por la primacía de la fuerza en las relaciones internacionales. Sus prácticas se asemejan a los patrones históricos del fascismo del siglo XX, con muchas similitudes en la lógica de poder y dominación. Así tenemos la captura del Presidente Maduro, en un operativo militar el 03 de enero 2026 (CNN, 2026), o el asesinato al líder supremo de Irán, Alí Jamenei, el 2 de marzo 2026 (BBC, 2026b), operaciones que buscan hacerse con el control de los recursos naturales, disfrazadas de exportación de la “democracia”.

Con relación a Medio Oriente, Trump presentó en el Foro Económico Mundial de Davos 2026, un plan para reconstruir la Franja de Gaza como una “nueva Dubái”, orientada a inversiones, turismo y desarrollo inmobiliario. La propuesta busca transformar un enclave cultural e histórico, en un centro económico, con rascacielos, puerto, aeropuerto y zonas tecnológicas. Sin embargo, mientras se presentaba esta visión de desarrollo económico, gran parte de la población de Gaza seguía enfrentando una grave crisis

humanitaria, tras años de bombardeos e incursiones militares violentas por parte de Israel (El País, 2026).

Esta iniciativa ha generado un gran escepticismo internacional, en vista de su desconexión con la realidad política y humanitaria del territorio. Podemos decir que se trata de un plan macabro, que prioriza una reconstrucción económica, sobre las necesidades reales del pueblo palestino, brutalmente sacrificado por la complicidad entre Israel y Estados Unidos (El País, 2026).

En el caso de Javier Milei en Argentina, una vez en el poder, no ha dudado en golpear directamente las conquistas sociales acumuladas en el tiempo. Su llegada a la Casa Rosada aceleró la política de ajuste fiscal, registrando una caída real de la inversión pública de 7,6 % interanual y de 6,1 % en el primer bimestre de 2026, con recortes especialmente significativos en programas sociales (cerca del 44,8 %), transferencias a provincias y gasto en personal estatal (Infobae, 2026). De igual forma ha impulsado una reforma laboral que permite flexibilizar las reglas del mercado y ampliar la jornada laboral diaria hasta 12 horas sin pagar horas extras (BBC, 2026a).

Esta reforma incluye la modificación del cálculo de indemnizaciones, reduciendo el costo de despido para las empresas. Incluye beneficios fiscales para empleadores que contraten trabajadores y facilita el pago de salarios en

moneda extranjera. El gobierno la presenta como una modernización para crear empleo, mientras sindicatos y opositores denuncian una reducción de derechos laborales (BBC, 2026a).

Masivas y constantes movilizaciones en Argentina reflejan la creciente confrontación social con el gobierno de Milei, que denuncian las afectaciones del ajuste económico y el desmantelamiento de políticas públicas. Las reformas laborales y recortes sociales implican precarización y aumento de la desigualdad. Estas dinámicas anticipan escenarios de mayor conflictividad interna, que pueden influir en la estabilidad política y en la legitimidad del programa económico del gobierno.

Por su parte, Daniel Noboa, tras su reelección en 2025, consolidó un modelo de gobierno cívico-militar, que emplea la narrativa de la “guerra interna” y la crisis de seguridad para legitimar un programa de ajuste económico vinculado al acuerdo con el Fondo Monetario Internacional, que incluye el aumento del IVA mediante el Decreto 198, la eliminación de subsidios y reformas económicas orientadas al mercado. La seguridad funciona como ideología de gobierno y mecanismo de control social, en una tendencia regional de “neoliberalismo punitivo” comparable con los modelos de Nayib Bukele y Javier Milei (Stoessel, 2025).

A pesar de esta propaganda securitista, Ecuador se ubicó como el sexto país más violento del mundo en 2025. En los dos primeros años del gobierno de Daniel Noboa, Ecuador registró 15.561 muertes violentas, lo que equivale a un asesinato cada 66 minutos. El promedio nacional alcanza alrededor de 22 homicidios diarios, reflejando la persistencia de una grave crisis de seguridad. A pesar de las políticas de militarización, estados de excepción y la declaración de conflicto armado interno, la violencia no ha logrado ser contenida. Se estima que el 71% de la población convive de forma directa con la violencia criminal, una de las cifras más altas de la región. Estos datos reflejan el deterioro de la seguridad en Ecuador y la consolidación de una crisis estructural: económica, social, de violencia y crimen organizado (Ecuavisa 2026).

En línea con la reducción y afectación de derechos, el 20 de febrero de 2026 la Asamblea Nacional aprobó la Ley de Sostenibilidad y Eficacia del Gasto de los Gobiernos Autónomos Descentralizados (GAD), promovida por Daniel Noboa. Reforma encaminada a limitar la capacidad de los municipios para financiar servicios esenciales para la ciudadanía. La iniciativa introduce una regla que restringe el gasto municipal en lo que el proyecto denomina “gasto no prioritario”, lo cual amenaza el financiamiento del personal que sostiene servicios públicos como salud, educación, programas sociales y transporte municipal. La reforma está encaminada a producir despidos masivos de servidores

municipales y a reducir o privatizar servicios públicos (Quito Informa, 2026).

Alcaldes y autoridades locales consideran que la medida vulnera la autonomía de los gobiernos autónomos descentralizados, al imponer criterios de gasto desde el Gobierno central. Advierten posibles problemas de constitucionalidad por cambios presupuestarios con efectos retroactivos y un gran impacto a los servicios públicos que reciben los ciudadanos (Quito Informa, 2026).

Como podemos apreciar, se trata de gobiernos que lideran proyectos políticos que no se contentan únicamente con administrar el orden existente, sino que buscan reconfigurar radicalmente la sociedad, las instituciones y las subjetividades. Lo hacen no solo a partir de las políticas de ajuste y privatización, sino también, apoyados en estrategias encaminadas a erosionar la solidaridad, desactivar la empatía y sembrar el miedo como afecto político central.

El objetivo del neofascismo contemporáneo, en palabras de Safatle (2025a), es doble: por un lado, destruir la posibilidad de un “nosotros” colectivo, que sustente lo principios de la justicia social, y, por otro lado, posicionar una cultura política basada en el antagonismo, la fragmentación y la guerra de todos contra todos.

Lejos de apelar a una masa unificada y fervorosa, como lo hacía el fascismo clásico, los gobiernos neofascistas de actualidad activan emociones disgregadoras, haciendo del "otro": el migrante, el feminismo, el indígena, el disidente, el chivo expiatorio que legitima sus activaciones y movilizaciones políticas. Este proceso de transformación autoritaria, no se opone al neoliberalismo, por el contrario, surge como su consecuencia y evolución lógica.

Si bien el neoliberalismo disuelve los vínculos sociales y empuja al sujeto hacia la autoexplotación, el neofascismo aparece como nuevo autoritarismo, que se legitima bajo el discurso de la libertad. Pero no se trata de la libertad como emancipación o derecho colectivo, sino, como señala Safatle (2025b), de una libertad regresiva, entendida como inmunidad frente a los demás, como rechazo al cuidado, a lo público y a la redistribución.

En este marco, la salud, la educación, la seguridad y la justicia dejan de ser derechos universales para convertirse en servicios privatizados y mercantilizados, accesibles solo a quienes pueden pagarlos. Para el neofascismo el Estado social debe ser desmantelado, y en su lugar se deben colocar pulsiones violentas como: racismo, individualismo, competencia despiadada; eliminando cualquier sentido de lo común, o las bases materiales de la democracia (Safatle, 2025b).

La pobreza, el desempleo, la desigualdad y la violencia no son combatidos en sus causas, sino que pretenden ser combatidos a partir de discursos de odio, militarización y punitivismo. Se impone así, como señala Safatle (2025a), una política del abandono, que normaliza la precariedad, criminaliza la diferencia y legitima el sufrimiento como destino natural.

El neofascismo es una corriente en auge que se articula globalmente. Como lo ha manifestado Safatle (2025b), los lazos entre presidentes como Bolsonaro, Bukele, Milei o Noboa, se consolidan en redes internacionales conservadoras, conferencias, *think tanks* y plataformas digitales. Se trata de un entramado que da lugar a una “internacional neofascista”, que opera ante todo en el campo de las emociones y de la guerra cultural.

Gracias a sus grandes contingentes económicos, la política puede ser trasladada de las calles a las redes, en donde las *fake news*, los discursos de odio y las campañas de desinformación reemplazan al debate racional y al disenso democrático (Safatle, 2025b).

Lo que estas plataformas promueven, no es simplemente una ideología, sino una ingeniería afectiva de masas, en la cual los algoritmos privilegian contenidos que provocan miedo, indignación o rechazo. Las redes sociales en este sentido, no son neutrales, puesto que amplifican los



discursos más extremos, invisibilizan los matices y castigan las voces que promueven la inclusión, el feminismo, el ecologismo o la justicia social.

Esto genera un escenario y un entorno que vacía de sentido a las democracias. Los ciudadanos quedan reducidos a meros consumidores de emociones, convertidos en instrumentos de proyectos autoritarios que niegan su propia autonomía. El neofascismo es un síntoma de la radicalización neoliberal, su mutación autoritaria busca por la fuerza, (del capital y el control de las ideas), estabilizar un orden profundamente injusto, a través del miedo, la división y el odio.

### *Tecnofeudalismo o la privación de la democracia*

La democracia entendida como un espacio de deliberación pública y construcción colectiva de sentido (Habermas, 1962), enfrenta hoy una amenaza inédita: a más del avance de regímenes neo-autoritarios, surge una forma más sutil y compleja de dominación, mediada por algoritmos, pantallas y flujos de datos. Como advierte Franco "Bifo" Berardi (2019), asistimos al retorno del fascismo en donde la violencia no se expresa abiertamente en el espacio público, sino que se disuelve en los circuitos de la información, transformando la subjetividad, desactivando la resistencia y gobernando los cuerpos y las mentes a través de la estimulación info-nerviosa.

Esta transformación marca el paso del neoliberalismo clásico hacia un nuevo régimen de poder: el tecnofeudalismo, término acuñado por Yanis Varoufakis (2024), para describir la emergencia de una clase dominante compuesta por los dueños de los grandes latifundios digitales: Amazon, Google, Facebook, Apple; quienes controlan la infraestructura tecnológica que estructura la vida social en el planeta.

El tecnofeudalismo constituye una mutación del capitalismo contemporáneo que favorece el surgimiento de dinámicas neofascistas, ya que las grandes plataformas tecnológicas han consolidado un poder económico y digital, que crea relaciones de dependencia similares a estructuras feudales (Varoufakis, 2024). El control sobre datos, algoritmos y flujos de información otorga a las corporaciones tecnológicas una enorme capacidad para influir en la opinión pública y en los procesos políticos. Enzo Traverso (2019), nos recuerda que las nuevas derechas radicales se sirven de estos ecosistemas digitales, para difundir y posicionar discursos nacionalistas, excluyentes y autoritarios, debilitando de esta manera las instituciones democráticas y los espacios públicos de deliberación.

Esta erosión institucional crea condiciones propicias para el ascenso de proyectos políticos antidemocráticos, que, en palabras del filósofo Vladimir Safatle (2025b), intensifican la polarización social y normalizan la violencia práctica y simbólica en contra de grupos vulnerables. Se trata en

nuestros días de tecnologías que están modelando el campo de la política, para abrir paso a propuestas autoritarias.

Ya no se trata solamente de capital financiero, sino de capital algorítmico, capaz de predecir, moldear y dirigir el comportamiento humano sin necesidad de coerción física, sin deliberación pública, sin legitimidad democrática. El espacio político, entendido como lugar de conflicto, deliberación y cambio (Habermas, 1962), ha sido clausurado por un ecosistema de plataformas que priorizan la rentabilidad por sobre el diálogo.

En lugar de espacios de conversación y deliberación entre los ciudadanos, se impone en nuestros días, una relación unidireccional entre el individuo y la pantalla, donde la información ya no circula libremente, sino que es jerarquizada, sesgada, amplificadas u ocultada según los intereses comerciales de las corporaciones digitales.

De esta forma el poder algorítmico suplanta la soberanía democrática, operando sin transparencia, sin regulación y sin responsabilidad política. Como bien señala Berardi (2019), este nuevo autoritarismo ya no requiere uniformar ideológicamente a la sociedad: le basta con fragmentarla, aislar a los sujetos, saturarlos de estímulos y polarizarlos emocionalmente.

La democracia, que presupone la posibilidad de un “nosotros” político, se ve corroída por una tecnología que individualiza la experiencia, intensifica el narcisismo digital y sustituye la acción colectiva por la performance individual. Como resultado tenemos una ciudadanía despolitizada, atrapada en burbujas de información, expuesta a flujos de odio, miedo y resentimiento, fácilmente manipulables por discursos neofascistas (Vaoufakis, 2024).

En este contexto, la política se convierte en una guerra cultural librada en territorios digitales. Las plataformas digitales mediante sus algoritmos, privilegian los contenidos que generan mayor interacción emocional, indignación, furia, miedo, lo cual favorece la proliferación de discursos de odio y la construcción de enemigos internos: migrantes, feministas, ecologistas, pobres.

Esta lógica, como advierte Varoufakis (2024), no es neutra: alimenta a esta nueva forma de fascismo, que sustituye la razón por la emoción, la deliberación por el algoritmo, el bien común por el consumo individual. El tecnofeudalismo no solo privatiza los servicios públicos, sino que se toma y privatiza el espacio mental.

La relación entre ciudadanía y poder se torna opaca, automatizada, mediada por tecnologías sin rostro ni rendición de cuentas. Se erosiona la posibilidad misma de resistencia, porque se pierde el lenguaje común, el terreno

simbólico compartido, la experiencia colectiva del malestar (Berardi, 2019; Varoufakis, 2014). La explotación ya no se percibe como tal, la injusticia se vuelve invisible, disuelta en la velocidad y fragmentación de la infoesfera.

En este nuevo escenario, los Estados nacionales se revelan impotentes. Carecen de capacidad para regular a las plataformas, para fiscalizar sus algoritmos, para proteger los derechos de los ciudadanos en el entorno digital (Berardi, 2019; Varoufakis, 2014).

La soberanía popular queda subordinada a una gobernanza privada global, que opera según la lógica del lucro y no del bien común. Lo que está en juego no es solo la economía o la privacidad, sino la posibilidad misma de la democracia como proyecto histórico y horizonte político. El tecnofeudalismo, entonces, no es solo una evolución del capitalismo, es una mutación autoritaria del poder. En un mundo hiperconectado, el aislamiento es la nueva forma de castigo; la invisibilidad, la nueva forma de muerte social (Berardi, 2019; Varoufakis, 2014).

*Neoliberalismo y necropolítica: cuando el mercado decide quién vive y quién muere*

La relación entre neoliberalismo y necropolítica es muy clara, y se centra en cómo las políticas económicas del neoliberalismo, al priorizar el mercado sobre la vida,

producen activamente formas de abandono, exclusión y muerte social (Mbembe, 2011; Stiglitz, 2012). La necropolítica neoliberal se manifiesta a través de la reducción o la falta de inversión pública, deterioro o eliminación de los servicios públicos, discriminación, exclusión social y deshumanización como parte del juego de los más fuertes sobre los más débiles (Mbembe, 2011).

El neoliberalismo no solo produce y reproduce desigualdad, sino que legitima el abandono y la exposición a la muerte, de millones de ciudadanos, ya sea por omisión (desfinanciamiento de servicios públicos), o por acción directa (represión, encarcelamiento, militarización).

Lo que para el neoliberalismo representa una neutralidad económica es, en realidad, una forma de gobierno que distribuye desigualmente no solo los bienes, sino también la posibilidad misma de vivir. En este sentido, la necropolítica opera como una práctica extrema del neoliberalismo, al reafirmar que no todas las vidas importan ni merecen protección.

El neoliberalismo al proclamar la libertad individual como principio supremo, desmantela las condiciones materiales que hacen posible el acceso a servicios públicos y a una vida más digna a los sectores sociales menos favorecidos, haciendo de esta manera evidente el poder necropolítico de decidir quién merece protección (aquel que puede pagar) y

quién puede ser abandonado, explotado o expuesto a la muerte (quien no cuenta con recursos económicos) (Mbembe, 2011; Stiglitz, 2012).

El neoliberalismo y la necropolítica reorganizan la economía como gestión del descarte. Los cuerpos que no producen valor económico: los pobres, los migrantes, los racializados, los habitantes de calle y de periferias, son tratados como excedentes o residuos del capital. En esta línea se puede apreciar como la necropolítica no opera solo a través de la violencia directa, sino mediante el abandono estructural, la exclusión institucionalizada y la indiferencia planificada (Mbembe, 2011; Stiglitz, 2012).

Para la necropolítica neoliberal el individuo es responsable de su éxito o fracaso, por tanto, aquellos que no logran acceder afirmativamente a la sociedad de consumo “fallan” y son culpables de su propio sufrimiento. Se trata de una práctica política que legitima la indiferencia frente al sufrimiento ajeno y despolitiza la desigualdad, convirtiéndola en un dato más de la economía.

La necropolítica neoliberal vacía de contenido político a la ciudadanía, reduciendo a los sujetos a meras mercancías portadoras de capital humano cuyo valor depende de su capacidad de generar rentabilidad. Todos aquellos que no pueden sostenerse en el mercado, son expulsados de la comunidad política. Esto se puede apreciar brutalmente en

las fronteras, donde los migrantes son regulados no como sujetos de derecho, sino como flujos que deben ser controlados, ralentizados o detenidos según la demanda del capital global.

De esta manera, el neoliberalismo radicaliza la necropolítica, al exponer a la muerte a grandes sectores de la sociedad. Los cuerpos que no sirven al capital son controlados, detenidos o abandonados a la muerte en fronteras, cárceles o barrios marginales. Se trata de régimen económico y cultural en donde la vida pierde su valor intrínseco, y las políticas públicas dejan de proteger a las y los ciudadanos, para servir los intereses de rentabilidad.

Las políticas de austeridad, el achique del Estado, el austericidio empujan a los estados de excepción, a la militarización de los territorios marginalizados y al abandono estructural de ciertas poblaciones que pierden su derecho a una vida digna, quedando expuestas a la muerte, en nombre de la seguridad o del crecimiento económico.

Se trata de una forma de racismo, entendido como racismo estructural Mbembe (2011), ahí su vínculo con el fascismo, puesto que se convierte en un mecanismo central para justificar quién resulta prescindible y quien no, se reproduce el principio de la división entre “nosotros” y “ellos” que legitima la violencia, la exclusión, la marginación y en



algunos casos la muerte selectiva (de los más afectados económicamente).

## **¿Cómo disputar la democracia frente al despojo neoliberal?**

El neoliberalismo no es simplemente un modelo económico que presenta fallas técnicas o desequilibrios coyunturales. Se trata de un proyecto político-ideológico que penetra en la arquitectura misma de la vida en común, erosionando las condiciones de posibilidad de una vida digna, basada en los principios de solidaridad y equidad. Socava el papel del Estado y la justicia redistributiva, amenazando directamente los cimientos de la democracia (Stiglitz, 2012).

La captura del Estado por parte de los grupos económicos vinculados al gran capital, transforma la política en un dispositivo de acumulación, a costa del sufrimiento de las mayorías. El ciudadano se degrada en consumidor, el trabajador en mercancía, y lo común en un botín a ser privatizado. De este modo, la promesa moderna de igualdad política se ve desplazada por un régimen de desigualdad estructural, legitimada como resultado natural del libre mercado (Stiglitz, 2012).

Frente a la grave crisis económica, política y social que deja la necropolítica neoliberal a su paso, la disputa por recuperar una democracia plena demanda una operación

doble: política y cultural. Política, porque exige redistribuir ingresos y recursos, reposicionando lo público como garante de la dignidad para las grandes mayorías. Cultural, porque implica disputar el imaginario neoliberal que reduce el valor de las personas a su capacidad de generar rentabilidad.

En este sentido, disputar la democracia es también disputar el sentido de lo humano, reafirmando que no todo en la vida puede ser mercancía u objeto de cálculo económico. La gran apuesta del siglo XXI no es únicamente frenar la voracidad del neoliberalismo y su necropolítica, sino recuperar la noción republicana del bien común, capaz de poner la economía al servicio de la vida y no al revés (Stiglitz, 2012).

Recuperar la democracia es activar un campo de lucha permanente contra toda forma de dominación que degrade la vida. En tiempos de neoliberalismo hegemónico, esta lucha se convierte en una tarea urgente y decisiva para devolver a la política su capacidad de construir un nosotros basado en solidaridad, justicia e igualdad.

### *El estado como regulador y garante del bien común*

Desde finales del siglo XX, las élites económicas y mediáticas han desplegado un trabajo persistente de ingeniería ideológica destinado a vaciar de legitimidad al Estado. La narrativa hegemónica -difundida en los años

ochenta y noventa aún hoy vigente- presenta al aparato público como un ente inherentemente ineficiente, corrupto y obstaculizador del “progreso”.

Frente a esta visión empobrecida y mercantilizada de lo público, Joseph Stiglitz (2002) sostiene que la recuperación del Estado como actor central de la vida económica y social no es una opción, sino una condición imprescindible para recomponer las bases de sociedades más justas. Un Estado democrático, fuerte y transparente no se limita a “corregir” las fallas del mercado, se erige como garante del interés general, protegiendo los bienes públicos y comunes, defendiendo los derechos de las mayorías frente a la voracidad de minorías concentradoras de poder económico.

Revalorizar el papel del Estado implica dotarlo de capacidad para intervenir en la estructura misma de la economía. Significa concebirlo como constructor activo de derechos universales -salud, educación, vivienda, trabajo digno, protección ambiental- para que no dependa de la buena voluntad del azar del mercado, ni del poder adquisitivo individual, sino de una voluntad política orientada hacia la justicia social. El Estado, en esta concepción republicana, no es un mero árbitro neutral, es un agente protagónico y político que organiza la convivencia en función del bien común y del interés general.

El desafío del presente siglo no radica en abolir el Estado en nombre de una “eficiencia” que solo favorece a unos pocos, sino en restaurar la confianza colectiva en lo público a través de políticas redistributivas, incluyentes y planificadas, capaces de recomponer el lazo social roto por décadas de mercantilización de la vida.

El Estado debe retornar a la escena central del debate público, como garante de una democracia sustantiva, capaz de convertir la igualdad en una realidad material y no solo en una consigna formal. No se trata únicamente de administrar recursos, sino de reinstalar el sentido de lo colectivo, para devolver a la ciudadanía la certeza de que el bien común no es un residuo del mercado, sino la razón misma de la institucionalidad y la política.

### *Redistribuir la riqueza para reconstruir la cohesión social*

Stiglitz en su obra: “*El precio de la desigualdad*” (2012) desmonta la narrativa del “mérito individual” como justificación del orden desigual. Señala que lo que se presenta como éxito económico personal forma parte en realidad del privilegio heredado o la extracción de rentas desde posiciones de poder estructural (capital, social, económico o simbólico Bourdieu, 1986).

Hemos creado un sistema, señala Stiglitz (2012), que recompensa el rentismo y penaliza la producción, Las

políticas redistributivas orientadas desde el Estado, no deben ser percibidas como una cuestión de compasión o caridad, sino como una condición de posibilidad para la justicia y la estabilidad democrática.

Las sociedades más igualitarias tienden a funcionar mejor, ya que hay un menor grado de violencia, menos enfermedades, mayor inversión en educación, en ciencia, tecnología e innovación. No es casualidad, no dice Mazzucatto (2022), que los países más desarrollados y competitivos del mundo, sean los que tienen una mayor inversión social y presencia pública.

Las políticas redistributivas reducen los costos que genera la exclusión, la marginación y el abandono, permitiendo que más personas participen productivamente del mercado. Cuando existe un Estado que apuesta con fuerza a la educación, salud y servicios públicos universales, libera capacidades individuales para poder innovar, crear empresas, y participar en la vida económica con un menor riesgo.

Una sociedad más justa es a su vez una sociedad más competitiva, porque permite que más ciudadanos desarrollen sus capacidades y participen activamente en el proceso económico. La expansión de la libertad individual, nos recuerda Mazzucatto (2022), solo florece allí donde se eliminan las de grandes fuentes de privación como: la pobreza, la violencia, la mala salud. la exclusión, etc. Una

economía que deja a amplios sectores de la población fuera de estos mínimos existenciales para poder llevar adelante una vida digna, no es plenamente funcional, ni políticamente viable o legítima.

Los datos recientes del *Informe sobre Desarrollo Humano 2023-2024* del PNUD confirman este diagnóstico: países como Suecia, Finlandia o Alemania, que encabezan los índices de competitividad y desarrollo humano, comparten un denominador común: fuerte presencia pública, baja desigualdad y una cultura política orientada al bien común. Por lo tanto, redistribuir la riqueza, lejos de ser una concesión populista, es un acto de restitución de la dignidad social y una precondition del desarrollo entendido como proyecto nacional y no como privilegio corporativo.

En este horizonte, Thomas Piketty (2014) plantea la urgencia de una reforma tributaria global basada en la progresividad fiscal como principio de justicia. Esto implica que los ciudadanos y empresas contribuyan al financiamiento del Estado de acuerdo con su capacidad real de pago. Para ello plantea la necesidad de instaurar un impuesto progresivo sobre la renta, el patrimonio y las herencias, así como un impuesto mundial sobre el capital, con el objetivo de desactivar la acumulación desmedida de riqueza.

La fiscalidad progresiva (más tiene más aporta) no debe entenderse únicamente como un instrumento técnico de

financiación del Estado, sino como una herramienta política de redistribución del poder y de construcción de una sociedad más solidaria e igualitaria. Únicamente a través de la institución de sistemas fiscales justos, se torna viable financiar los derechos universales como salud, educación, vivienda y protección social, garantizando que la riqueza producida socialmente beneficie a todos y no solo a una minoría.

Se trata de una propuesta que hace posible revertir las lógicas del neoliberalismo mediante un sistema impositivo progresivo, que permite consolidar el contrato social y reforzar la legitimidad de las instituciones democráticas, reduciendo el poder desmedido del capital nacional y transnacional.

### *Acceso a conocimiento y educación como pilares centrales de la democracia*

En la cartografía contemporánea de las desigualdades, la mirada suele concentrarse en la distribución de la riqueza o de los ingresos. Sin embargo, como advierte Thomas Piketty (2019), existe una capa más profunda y determinante: la desigualdad en el acceso al conocimiento, a la cultura y al capital educativo. Este desnivel cognitivo no es un mero efecto colateral del mercado, sino un mecanismo estructural que reproduce jerarquías y consolida el poder de las élites.

La educación, según Piketty (2019), no es simplemente un medio para adquirir competencias laborales, sino una institución política clave para formar ciudadanos capaces de participar críticamente en la vida democrática. Es por ello que sugiere, sobre todo en los países emergentes, llevar a cabo una inversión masiva en educación pública, gratuita, universal y de calidad, con miras a eliminar las brechas territoriales, económicas y culturales entre los estudiantes.

Sin embargo, bajo el paradigma neoliberal, la educación pública ha sido objeto de un doble proceso: desfinanciamiento y mercantilización. La escuela deja de ser un espacio común y se fragmenta en circuitos segmentados por clase social, territorio y capital cultural. Este sistema segmentado asegura a las clases dominantes no solo la transmisión hereditaria de sus privilegios materiales, sino también el monopolio de las herramientas intelectuales con las que se interpreta y se interviene el mundo (Piketty, 2019).

La democratización del conocimiento resulta entonces, no solo un acto de justicia social, sino también de emancipación política necesaria. Una democracia sin ciudadanos formados, informados y críticos, es un cascarón vacío, incapaz de sostener instituciones inclusivas o resistir las derivas autoritarias que se incuban en la desigualdad. Por ello, Piketty (2019) insiste en que la lucha contra la desigualdad debe ser también una lucha contra la



concentración del saber y contra los dispositivos ideológicos que justifican la injusticia como destino inevitable.

La propuesta educativa, como señala Piketty (2019), apunta a transformar la estructura misma donde se inician las desigualdades, haciendo del conocimiento un bien común y no un privilegio de clase, elemento indispensable para poder avanzar hacia una sociedad en la que todos y todas tengan no solo derecho a la educación, sino derecho al poder que el conocimiento otorga. La política educativa se revela como una política del poder: o se distribuye para emancipar, o se concentra para dominar y excluir.

*Repolitizar la economía para desmontar el régimen ideológico neoliberal*

El neoliberalismo no debe entenderse como una doctrina técnica o una verdad económica objetiva, sino como una construcción ideológica profundamente política, diseñada para proteger y legitimar los intereses de las élites económicas. Plantean hacer de la austeridad, la privatización o la flexibilización laboral, verdades incuestionables. Sin embargo, como advierte Piketty (2019), se trata de una “ideología propietarista”, que pretende naturalizar la concentración de riqueza, debilitar la acción colectiva y culpabilizar a los pobres por su situación.

Esta operación ideológica no es inocente: busca fijar en la cultura política la idea de que “no hay alternativa” -el famoso *there is no alternative* de Margaret Thatcher-, clausurando el debate democrático sobre otras posibles formas de organización de la economía. Para poder superar al neoliberalismo, se torna indispensable desideologizar la economía, es decir, reconocer que las instituciones fiscales, laborales, educativas o financieras pueden organizarse de formas diferentes, con un enfoque mayor en la distribución de la riqueza, en el diseño de sistemas tributarios equitativos, así como en la garantía de derechos sociales universales.

En este sentido trabajar para profundizar y expandir la democracia implica una tarea doble, desideologizar y repolitizar la economía, para devolverle a la ciudadanía el poder de decidir cómo quiere vivir. Como podemos apreciar, la lucha contra la desigualdad no es solo una cuestión de redistribución económica, sino también, la necesidad de llevar a cabo una transformación cultural e ideológica. Hay que desmontar el discurso que presenta la desigualdad como inevitable y reactivar una nueva narrativa donde primen los principios de solidaridad y equidad como pilar para la construcción y fortalecimiento de instituciones volcadas al servicio de la gente y no del capital.

Repolitizar la economía no es entonces, un acto técnico, sino, un acto profundamente democrático: significa devolver a la ciudadanía el poder de decidir qué tipo de

economía y de sociedad quiere construir, es una apuesta por devolver a la política su lugar como potencia transformadora.

*Desmercantilizar la vida social, reconstruir lo común*

El neoliberalismo, nos dice David Harvey (2015), tiende y busca mercantilizar todas las dimensiones de la vida humana. Se trata de un modelo de acumulación por desposesión, que pretende privatizar la educación, la salud, el conocimiento e incluso la naturaleza, lo que implica que el acceso a los bienes esenciales ya no depende del reconocimiento de derechos, sino de la capacidad de pago lo cual termina profundizando las desigualdades históricas y estructurales.

Esta mercantilización integral de la existencia no solo redefine las relaciones económicas, sino que reconfigura la cultura política: allí donde antes había derechos universales, aparecen contratos; donde había bienes comunes, surgen mercados; donde había cuidado colectivo, impera la lógica del lucro. El Estado, lejos de actuar como garante del bienestar de la sociedad, es reorientado hacia la protección de los intereses del capital, legitimando así la necropolítica que termina decidiendo qué vidas son dignas de ser vividas y cuáles pueden ser descartadas.

Esto permite comprender por qué los sistemas públicos son desfinanciados o privatizados, y por qué se

activan campañas de empresas privadas que invitan a los individuos a “invertir” en educación, salud o seguridad (privadas), en lugar de proteger y garantizar los derechos ciudadanos. El neoliberalismo y su necropolítica ponen al Estado a proteger los intereses del capital.

Disputar la democracia y el acceso a derechos implica necesariamente, revertir los principios –ideológicos- de mercantilización de la vida social, a través de políticas públicas fuertes, servicios universales y la recuperación de los bienes comunes. Esto implica reconstruir lo público como espacio necesario de solidaridad, acceso equitativo y participación democrática, dimensiones de la vida social que no pueden ser delegadas al mercado o a entidades privadas motivadas únicamente por el beneficio económico.

El cuidado, el conocimiento, la salud, las jubilaciones, el ambiente, el tiempo, deben ser recuperados como bienes comunes, al ser indispensables para poder llevar una vida digna. Reposicionar la función del Estado quiere decir, reactivarlo como garante del bienestar colectivo y de la justicia social. La verdadera libertad no puede existir en medio de la escasez, la penuria, la violencia o de grandes desigualdades. Los derechos sociales no son mercancías, no pueden depender de la capacidad de consumo (Harvey, 2015).

La democracia sustantiva se enlaza necesariamente con la desmercantilización de los campos de la vida que

permite a los ciudadanos vivir con dignidad, autonomía y seguridad, sin tener que verse subordinados a las reglas impuestas por los mercados. Desmercantilizar la vida social pasa a ser hoy, un acto de resistencia frente al neoliberalismo y la necropolítica; una apuesta afirmativa por otro modelo de sociedad, donde la vida, el cuidado y la empatía -y ya no el capital-, pasen a ocupar el centro de las preocupaciones políticas.

*Hacia una democracia postneoliberal basada en la justicia, la cooperación y la sostenibilidad*

No obstante, el problema de las sociedades neoliberales y necropolíticas no radica únicamente en la desregulación, la privatización o la mercantilización de los diferentes campos de la vida, sino también, en su lógica expansiva, acumulativa y explotadora que pretende definir al capital como forma de organización social. Esta pretensión totalizante busca naturalizar la competencia, el individualismo y la precariedad, inscribiéndolos en los cuerpos, los afectos y las subjetividades.

Por ello, Harvey (2015) ha señalado con insistencia que la tarea histórica no consiste en maquillar el sistema con reformas cosméticas, sino en imaginar y construir una transición democrática capaz de subvertir los fundamentos del orden neoliberal. Esta transición implica una ruptura epistemológica y material: reemplazar la competencia por la

cooperación, el dominio corporativo por el control democrático de los medios de producción, la acumulación ilimitada por la justicia social y redistribución radical, y el productivismo ciego por una sostenibilidad ecológica que reconozca los límites planetarios y el valor intrínseco de los ecosistemas.

Pero esta transformación no emergerá de forma espontánea. Requiere la activación de un sujeto colectivo que desborde la pasividad ciudadana y organice la potencia de movimientos sociales, sindicatos, comunidades indígenas, colectivos feministas, redes de apoyo mutuo y ambientalistas. Retomando a Safatle (2016), podemos decir que para frenar o derrotar al neoliberalismo y su necropolítica, es necesario no solo redistribuir recursos materiales, sino también, ser capaces de generar nuevas formas de subjetividad, de sensibilidad y de vínculos, donde las personas puedan relacionarse más allá del miedo, del rendimiento o del castigo.

En un mundo marcado por la desigualdad extrema, la precarización de la vida y el colapso climático, superar al neoliberalismo y su necropolítica se torna una condición de supervivencia humana, ecológica y democrática. La transición hacia una democracia más justa e incluyente debe abrir espacios para redefinir las relaciones sociales, las formas de propiedad, la noción de trabajo y los valores culturales que permitan poner en valor y proteger lo común, redistribuir poder y facilitar formas de vida postneoliberales.

Cabe tener presente, como sostiene Safatle (2016), que el poder opresor y excluyente del neoliberalismo no viene solo de afuera, de las instituciones, y del gran sector del capital, sino que se produce también desde dentro del sujeto, en sus miedos, en su autocensura, en la incapacidad de imaginar que otra vida es posible. De ahí la importancia de generar formas colectivas de subjetivación, donde los individuos se constituyan no desde la competencia, sino desde el vínculo, el cuidado mutuo y la potencia de lo común.

Aquí cobra valor político experiencias como el feminismo, los movimientos antirracistas, las comunas, o las redes de apoyo mutuo, todas estas experiencias de activación y movilización de formas de vida capaces de reinventar la solidaridad como experiencia sensible, como reconfiguración radical de la subjetividad política, reapropiación colectiva de otras alternativas políticas posibles, capaces de romper con el realismo despiadado que nos impone el neoliberalismo como único horizonte.

Frente al neoliberalismo y la ultraderecha, el desafío, nos dice Cadahia (2025), es encontrar lazos que nos unan como seres humanos. La verdadera clave para la emancipación consiste en repolitizar la democracia, devolverle su dimensión conflictiva, afectiva y transformadora. Asumir que la política no es solo un juego de reglas fijas e inamovibles, sino un campo de disputa simbólica

y material, donde se pueden construir nuevos sentidos de lo común.

Superar el neoliberalismo y su necropolítica no es solo una tarea económica, es una condición de supervivencia humana, ecológica y democrática. Implica redefinir las relaciones sociales, las formas de propiedad, la noción misma de trabajo y los valores que orientan la vida colectiva. Es en última instancia, una lucha por reapropiarnos de nuestra capacidad de decidir, de imaginar y de vivir más allá del horizonte estrecho que el capital pretende imponernos como destino inevitable (Cadahia, 2025; Safatle, 2016).

## **Conclusión**

Una vez revisados los diferentes acápites que conforman este trabajo, podemos decir que la desigualdad y con ella, la pobreza y la violencia, nos son un accidente, ni un fenómeno natural inevitable. Por el contrario, son una consecuencia de decisiones políticas orientadas por principios económicos, que forman parte del proyecto político neoliberal orientado a distribuir el poder y los recursos en función de los intereses particulares y corporativos, principalmente de las élites económicas.

El ataque permanente al Estado y la reducción-anulación de sus funciones a mero facilitador de las leyes del mercado, así como la mercantilización de todas las



dimensiones esenciales de la vida humana, como: salud, educación, seguridad social, medio ambiente, etc., han provocado una fractura profunda en el tejido social, produciendo y reproduciendo exclusión sistemática y pobreza estructural.

Fenómenos como el neofascismo, la necropolítica, o el tecnofeudalismo, deben entenderse hoy, como extensiones autoritarias y violentas que configuran y forman parte del neoliberalismo, y que están designadas a garantizar la continuidad y permanencia del orden neoliberal, frente a cualquier tensión social que éste pueda generar.

La desprotección deliberada de los sectores vulnerables, gente en situación de calle, de escasos recursos, migrantes, minorías; evidencia una gestión política que decide quién merece vivir y quién queda expuesto a condiciones infrahumanas o al abandono sistémico, por parte de un Estado funcional a los intereses del capital,

Frente a estas mutaciones autoritarias y excluyentes del neoliberalismo, se vuelve imperativo replantear y reposicionar el rol del Estado en la construcción de sociedades más justas y democráticas. Revalorizar o reconocer el papel central que puede jugar un Estado activo, solidario y comprometido con el desarrollo de todos los sectores de la sociedad (sin exclusión), implica rediseñar sus funciones para garantizar derechos fundamentales y condiciones materiales

mínimas, que permitan la participación plena e igualitaria en la vida política y social.

Un Estado social fuerte en términos democráticos, debe reconstituir su capacidad reguladora para limitar los excesos del mercado y corregir las desigualdades estructurales que este genera. Está en la capacidad de promover políticas públicas orientadas a la inclusión social, al acceso universal a servicios básicos y a la protección efectiva de los derechos humanos.

La recuperación estatal viene empujada y acompañada por el fortalecimiento de los procesos democráticos, la activación y la movilización social, que van mucho más allá de la mera representación formal a elecciones cada 4 o 5 años. Solo desde un Estado comprometido con las necesidades reales de las grandes mayorías, se podrá construir una democracia capaz de enfrentar las múltiples crisis sociales y económicas derivadas del neoliberalismo.

Los tiempos que corren ahora en países como Argentina (Milei), Estados Unidos (Trump) o Ecuador (Noboa), son de mucho dolor y de muchas injusticias sociales. En este escenario, la política tiene una responsabilidad enorme, o mantiene el estado de las cosas, o las transforma para mejorarlas.

La política moderna, heredera de la revolución francesa, tiene una matriz revolucionaria, que es vital para poder mantenerse viva. Esto nos convoca como ciudadanos y organizaciones sociales, a poder denunciar las fallas de los sistemas excluyentes, evitar que las élites consoliden poderes autoritarios, y canalizar por vías democráticas, las demandas de los sectores excluidos.

En tiempos donde campea la violencia, el crimen y la desigualdad rampante, apalancadas por un sistema – neoliberal- excluyente, esta matriz –revolucionaria- cobra vida nuevamente, porque contiene en ella el motor social que impulsa los cambios profundos para lograr sociedades más justas, solidarias e incluyentes, capaces de superar las divisiones (ganadores – perdedores) impuestas por lógicas de mercado que generan opresión, exclusión y desigualdad.

La recuperación del Estado constituye una condición necesaria para garantizar no solo un crecimiento económico más equitativo, sino también, una democracia legítima, donde los sujetos ya no sean objetos descartables, sino que puedan recuperar su voz, así como los derechos efectivos y el acceso a una vida digna.

## **Referencias**

Berardi, F. (2019). "El problema es cómo la pantalla se ha apoderado del cerebro". *El País*.

[https://elpais.com/cultura/2019/02/18/actualidad/1550504419\\_263711.html](https://elpais.com/cultura/2019/02/18/actualidad/1550504419_263711.html)

- Bourdieu, P. (1986). The forms of capital. En J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood.
- Cadahia, L. (2025). "Un antídoto contra el fascismo". *Palabra Pública*. <https://palabrapublica.uchile.cl/un-antidoto-contra-el-fascismo/>
- Dubet, F. (2021). *¿Por qué preferimos la desigualdad? (Aunque digamos lo contrario)*. Siglo XXI.
- Echeverría, B. (2010). "América Latina: 200 años de fatalidad". *Revista Sin Permiso*. <https://www.sinpermiso.info/textos/amrica-latina-200-aos-de-fatalidad>
- Friedman, M. (1962). *Capitalism and freedom*. University of Chicago Press.
- Habermas, J. (1962). *Espace public et démocratie délibérative: Un tournant*. Gallimard.
- Han, B.-C. (2024). *La sociedad del cansancio*. Herder.
- Harvey, D. (2005). *Breve historia del neoliberalismo*. Akal.
- Harvey, D. (2015). *17 contradicciones y el fin del capitalismo*. IAEN.
- Hayek, F. A. (1944). *The road to serfdom*. George Routledge & Sons.
- Honneth, A. (1997). *La lucha por el reconocimiento: Por una gramática moral de los conflictos sociales*. Crítica.

- Mazzucato, M. (2022). *El Estado emprendedor: La oposición público-privado y sus mitos*. Taurus.
- Mbembe, A. (2011). *Necropolítica*. Melusina.
- Piketty, T. (2014). *El capital en el siglo XXI*. Fondo de Cultura Económica.
- Piketty, T. (2019). *Capital e ideología*. Deusto.
- Programa de las Naciones Unidas para el Desarrollo. (2024). *Informe sobre Desarrollo Humano 2023/2024: Romper el bloqueo: Reimaginar la cooperación en un mundo polarizado*. PNUD.  
<https://hdr.undp.org/content/human-development-report-2023-24>
- Safatle, V. (2016). *Grand Hotel Abyss: Desire, recognition, and the restoration of the subject (Figures of the Unconscious)*. Leuven University Press.
- Safatle, V. (2025a). "Bolsonaro se ve al frente de una revolución en marcha y no va a parar". *Revista Nodal*.  
<https://www.nodal.am/2020/05/vladimir-safatle-filosofo-bolsonaro-se-ve-al-frente-de-una-revolucion-en-marcha-y-no-va-a-parar/>
- Stiglitz, J. (2002). *El malestar en la globalización*. Taurus.
- Stiglitz, J. (2012). *El precio de la desigualdad: El 1% de la población tiene lo que el 99% necesita*. Taurus.
- Stoessel, S. (2025). "Guerra interna y paro nacional en Ecuador". *Jacobin*.

- Traverso, E. (2019). *Las nuevas caras de la derecha: Potencia y contradicciones de la etapa posfascista. Siglo XXI.*
- Varoufakis, Y. (2024). *El capitalismo ha muerto y el sistema que lo reemplaza no es mejor.* Deusto.
- Wolin, S. (2010). *Democracia S.A.: La democracia dirigida y el fantasma del totalitarismo invertido.* Katz Editores.

### **Videos:**

- Safatle, V. (2025b). *Por que o fascismo nos espera* [Video]. Opera Mundi.  
<https://www.youtube.com/watch?v=JjvOOGSVy3k>

### **Prensa:**

- BBC. (2026a). "3 claves de la polémica reforma laboral impulsada por Javier Milei y aprobada por el Senado de Argentina". *BBC News Mundo*.  
<https://www.bbc.com/mundo/articles/cq8g0y235wqo>
- BBC. (2026b). "Qué se sabe del operativo de EE.UU. e Israel con el que localizaron y mataron al líder supremo de Irán, Alí Jamenei". *BBC News Mundo*.  
<https://www.bbc.com/mundo/articles/c309j9mpmvi>
- CNN. (2026). "Así capturaron a Maduro: Análisis de CNN revela riesgos extremos del operativo militar en Venezuela". *CNN en Español*.

<https://cnnespanol.cnn.com/2026/01/23/venezuela/maduro-captura-riesgos-eeuu-trax>

Ecuavisa. (2026). "Ecuador se convirtió en el sexto país más violento del mundo". *Ecuavisa*.

<https://www.ecuavisa.com/seguridad/ecuador-sexto-pais-violento-mundo-20260216-0001.html>

El País. (2026). "El Gobierno de Trump vende en Davos una 'nueva Gaza' al estilo de Dubái". *El País*.

<https://elpais.com/internacional/2026-01-22/el-gobierno-de-trump-vende-en-davos-una-nueva-gaza-al-estilo-de-dubai.html>

Infobae. (2026). "Milei aceleró el ritmo del ajuste en febrero: Los gastos que más bajaron". *Infobae*.

<https://www.infobae.com/economia/2026/03/09/milei-acelero-el-ritmo-del-ajuste-en-febrero-los-gastos-que-mas-bajaron/>

Primicias. (2026). "Estados Unidos deportó a 742 ecuatorianos en los primeros 26 días de enero de 2026". *Primicias*.

<https://www.primicias.ec/sociedad/ecuatorianos-deportados-estados-unidos-enero2026-paso-darien-cancilleria-114701/>

Quito Informa. (2026). "'Defenderemos a las familias más necesitadas y sus servicios': Pabel Muñoz sobre aprobación de reforma al COOTAD". *Quito Informa*.

<https://www.quitoinforma.gob.ec/2026/02/21/defenderemos-a-las-familias-mas-necesitadas-y-sus->

[servicios-pabel-munoz-sobre-aprobacion-de-reforma-  
al-cootad/](#)



# Inteligencia artificial y tecnoceno: genealogía, control y dilemas éticos de la revolución algorítmica

*Artificial Intelligence and the Technocene: Genealogy, Control, and Ethical Dilemmas of the Algorithmic Revolution*

**Ana Sedeño-Valdellos<sup>40</sup>, Xavier Brito Alvarado<sup>41</sup>**

## Resumen

El presente ensayo analiza el impacto de la Inteligencia Artificial (IA) en la sociedad contemporánea, abordando sus implicaciones tecnopolíticas, éticas y culturales. Examina la reconfiguración de las relaciones de poder, los

---

<sup>40</sup> Universidad de Málaga. <https://orcid.org/0000-0003-3897-2457>

<sup>41</sup> Universidad Técnica de Ambato. <https://orcid.org/0000-0001-8593-3691>

cuestionamientos al humanismo y los riesgos de carácter existencial, y explora narrativas alternativas, como el aceleracionismo. La metodología se basa en la teoría crítica, recurriendo a autores como: Bostrom, Sadin, Floridi, Žižek y Land, y a marcos conceptuales de la filosofía, la sociología y los estudios de ciencia y tecnología. El trabajo se organiza en un recorrido histórico y conceptual de la IA, seguido de un análisis crítico de sus efectos normalizadores, su control algorítmico y los dilemas éticos que genera.

**Palabras clave:** Inteligencia artificial; ética; humanismo; algoritmos; tecnopolítica.

## **Abstract**

This essay analyzes the impact of Artificial Intelligence (AI) on contemporary society by addressing its technopolitical, ethical, and cultural implications. It examines the reconfiguration of power relations, the challenges to humanism, and existential risks, while also exploring alternative narratives, such as accelerationism. The methodology draws on critical theory, engaging with key authors including Bostrom, Sadin, Floridi, Žižek, and Land, as well as conceptual frameworks from philosophy, sociology, and science and technology studies. The work is structured around a historical and conceptual overview of AI, followed by a critical analysis of its normalizing effects, algorithmic control, and the ethical dilemmas it generates.

**Keywords:** Artificial Intelligence; Ethics; Humanism; Algorithms; Technopolitics.

## **Introducción**

La publicación de *El origen de las especies* (Darwin, 1859) impulsó un replanteamiento fundamental: la biología no representa el punto final de la evolución humana. Para Nick Bostrom (2011), esta obra hizo tangible la noción de que la evolución natural puede ser modificada mediante la tecnología. Jeremy Rifkin (2011) denomina al siglo XXI como el de la biotecnología, una revolución anclada en el presente donde el conocimiento se instala en una matriz operativa que, desde los años cincuenta, no ha cesado de ofrecer soluciones a los males humanos, aunque también ha abierto interrogantes y temores sobre sus fines.

Vivimos una época de incertidumbres y ausencia de certezas. El progreso científico ha sido puesto en entredicho, desafiando el proyecto humanista moderno. Con frecuencia, el conocimiento científico-técnico demuestra su incapacidad para reflexionar sobre su propia aceleración. Para Javier Echeverría (2003), los fundamentos de la ciencia surgidos en la temprana modernidad persisten, pero hoy cuentan con tres herederos, la microciencia, la macrociencia y la tecnociencia, que han reconfigurado el rumbo de la humanidad y contribuido a generar cambios en el orden económico, político y social. Esto ha creado imaginarios de miedos y

fantasías que colonizan los debates sociales, anclados en un misticismo terrorífico sobre el fin de la humanidad, que Herminio Martins (2018) evoca como una actitud fáustica orientada a trascender la condición humana. Este escenario refleja cómo el dominio de la racionalidad técnica ha conducido a una forma conectiva de nuestra existencia, enmarcada por la hiperconexión digital nos *expone al riesgo* de transformarnos en autómatas regidos por estímulos y respuestas.

Las posturas a favor o en contra del uso de estas tecnologías suelen inscribirse en la esfera del tecnorromanticismo, un entramado de tecnofilia, idealismo, temores y nostalgia por una sociedad tecnologizada. Este fenómeno presenta un nuevo ethos para la humanidad con una doble postura: la visión utópica, donde las tecnologías construirán sociedades justas y solidarias, y la distópica, que proyecta un futuro aterrador dominado por tiranos tecnológicos.

Marshall McLuhan (1996) argumentaba que, desde los años setenta, “todos somos robots cuando nos involucramos acríticamente con nuestras tecnologías”. Esta frase premonitoria se ha cumplido al generar una forma de sociabilidad basada en la conexión digital, que ha dado lugar a un humanismo tecnológico que exige una reconfiguración de la arquitectura neurobiológica del sujeto y la sociedad.

Para muchos científicos, la neurociencia y la neuroingeniería convergerán hasta constituir un solo campo. Los neurocientíficos podrán poner a prueba sus teorías mediante modelos cerebrales artificiales, lo que permitirá acelerar la comprensión de cómo surge y se encarna la inteligencia. Sin embargo:

lo que aún está en disputa son las maneras en que este cambio afecta la superposición entre conocimiento y poder, o, mejor dicho, cómo los cambios tecnológicos podrían incorporarse a una comprensión inherentemente política de la teoría contemporánea del conocimiento. (Land, 2014, s.p.)

Entre todas las tecnologías, la inteligencia artificial emerge como la de mayor desarrollo, hasta convertirse en una verdad incuestionable. La aparición de *large linguistic models* (LLM), como ChatGPT, Gemini, DeepSeek, TextCortex, entre cientos, ha configurado una nueva cotidianidad y geopolítica en el orden tecnológico, epistemológico, comercial, político, económico, moral y ético. La revolución de las máquinas digitales comienza a ocupar el lugar de la inteligencia y el razonamiento del individuo. Esta preocupación no es nueva: René Descartes, en el siglo XVII, reflexionaba sobre la razón imputada a la mente individual y la posibilidad de instrumentalizarla como cálculo puro.

Para autores como Éric Sadin (2020) y Herminio Martins (2018), la evolución humana estaría transitando simbólicamente del *Homo sapiens* al *Homo silicium*, y de ahí al *homo algorithmicus*, una figura que condensa la tesis de que la existencia humana comienza a ser regulada por la lógica algorítmica. Yoshua Bengio (2023) plantea preguntas fundamentales sobre estos cambios: ¿cómo deberíamos pensar la aparición de sistemas de inteligencia artificial (IA) incluso superhumanos? ¿Deberíamos abrazarlos por su potencial para mejorar nuestras vidas o temerlos por su capacidad de despojar a la humanidad de poder y, posiblemente, conducirla a la extinción?

Los sistemas modernos de IA ejecutan operaciones algorítmicas basadas en datos, lo que puede resultar en la discriminación de grupos marginados sin acceso a dicha tecnología. En este contexto han surgido subcampos como el estudio de la injerencia de la IA y la ética en los negocios, la educación y la toma de decisiones políticas. El uso de la IA implica un riesgo relevante para la democracia, así como la concentración de poder y riqueza en pocas manos debido al creciente poder de sus herramientas. Esto podría afectar a trabajadores, consumidores, la eficiencia de los mercados y la seguridad global, utilizando datos personales cedidos libremente en internet sin plena conciencia de las implicaciones. La concentración extrema podría otorgar un poder inédito, contradiciendo los principios democráticos y representando una amenaza significativa.

Se inaugura así una etapa que Martins (2018) denominó *Tecnoceno*, caracterizada por la llegada de tecnología de alta complejidad y riesgo, accesible a cualquier sujeto. Por su parte, Flavia Acosta (2024) argumenta que el mundo actual se forjó en las últimas cinco décadas, marcadas por una aceleración científico-técnica, biológica y social, cuyo resultado inmediato es la masificación de la IA en todos los ámbitos de la vida humana.

Para este abordaje el ensayo se estructura en de la siguiente manera. 1) se traza una genealogía conceptual de la IA, desde los autómatas de Turing hasta los modelos de lenguaje, 2) se examina la cultura del algoritmo, analizando cómo la IA opera como un nuevo poder de normalización que, a diferencia de la modernidad clásica, ya no reside en el Estado sino en las corporaciones tecnológicas; 3) se adentra en los dilemas éticos que esta tecnología plantea, confrontando las propuestas de autores como Floridi Bostrom o Žižek, y la 5) el aceleracionismo, no como una desviación del argumento, sino como su necesario horizonte ideológico: el aceleracionismo representa la respuesta filosófica y política más radical a la pregunta de si debemos regular la IA o, por el contrario, impulsar su desarrollo hasta sus últimas consecuencias.

## **La revolución algorítmica**

En el horizonte civilizatorio asoman conceptos como cataclismo, apocalipsis, colapso y aniquilación. Para Alex Williams y Nick Srnicek (2015), las revoluciones tecnológicas han modificado las experiencias políticas y sociales, condicionadas por dinámicas antropocénicas espacio-tiempo que sitúan en el centro sus efectos más dramáticos. Rosi Braidotti (2020) caracteriza esta época por la yuxtaposición entre la quinta revolución industrial (digital) y la sexta extinción masiva de vida en el planeta.

La IA representa la continuación de la civilización humana dentro de una tradición faustiana, donde se apuesta por un futuro determinado mediante una transformación radical del presente. El transhumanismo y el posthumanismo, temas recurrentes en los ámbitos intelectual, académico, económico, social y político, han recogido esta propuesta, aunque a menudo derivan hacia narrativas de ficción distópica que desplazan las advertencias científicas. La ontología de la IA se fundamenta en la noción liberal del sujeto autónomo planteada por Locke y Kant, basada en la capacidad individual para pensar, decidir, discernir entre el bien y el mal, y transformar las condiciones de vida.

En la década de 1930, Alan Turing y Emil Post postularon la construcción de autómatas capaces de realizar cálculos y determinar la validez de teoremas mediante la



manipulación de símbolos. En 1943, Konrad Zuse construyó una computadora capaz de emular la máquina de Turing, lo que constituyó la base de los lenguajes de programación. Ese mismo año, Warren McCulloch y Walter Pitts propusieron un modelo de computación alternativo basado en el funcionamiento de las neuronas biológicas: un teorema de lógica de primer orden determinado no por un autómata, sino por una red de neuronas artificiales que se excitaban mutuamente mediante conexiones sinápticas.

Turing (1950) propuso un experimento con tres participantes, un hombre (A), una mujer (B) y un interrogador (C), cuyo objetivo era que el interrogador identificara a cada uno mediante preguntas escritas. Turing hipotetizó que una máquina podría considerarse inteligente si lograba engañar al interrogador haciéndole creer que era humana. Este principio subyace a sistemas contemporáneos como el CAPTCHA, diseñado para distinguir entre usuarios humanos y máquinas.

En 1955, los organizadores de la Propuesta de Dartmouth argumentaban que “todo aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse con tal precisión que una máquina pueda ser construida para simularlo” (McCarthy et al., 2006, p. 15). En la misma década, Claude Shannon presentó la teoría de la información y desarrolló una máquina capaz de jugar al ajedrez mediante el método *minimax*, que evalúa los mejores

movimientos. En 1956, John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon acuñaron el término "inteligencia artificial" y establecieron sus bases: a) el pensamiento puede ocurrir fuera del cerebro; b) es posible una comprensión formal y científica del pensamiento; c) la inteligencia puede ser simulada mediante ordenadores.

Hubert Dreyfus (1965) sostuvo que la inteligencia humana tiene un límite en las redes neuronales. Sin embargo, esta perspectiva "no logra captar la esencia de los modelos de lenguaje de gran escala, los cuales constituyen un modelo estadístico del lenguaje humano extraído de la web" (Halpin, 2025, p. 4589). El individualismo ontológico es central en la IA, entendida como la artificialidad que asume una propiedad humana al crear un orden lógico a partir de la información. Por ejemplo, ChatGPT puede demostrar inteligencia al razonar y mantener una conversación. Desde una perspectiva ideológica, la IA parecería tender a perfeccionar un modelo de control: "El propósito de la IA no consiste únicamente en sustituir a los seres humanos, sino en administrarlos, preservando las relaciones de poder existentes" (Halpin, 2025, p. 4589).

La fuente de inteligencia de la IA no es autónoma; reside en una alternancia entre humanos y máquinas, por lo que representa una colaboración colectiva. Esto permite explicar cómo las computadoras tejen inteligencia humana y no humana, generando formas complejas de datos

compartidos. Su lógica es la creación de un razonamiento mecanizado similar al humano con la finalidad de capturarlo, una analogía que presupone a la IA como una forma análoga al sujeto ilustrado del racionalismo moderno.

Un hito ocurrió en 1973, cuando el *Center for Human Modeling and Simulation* de la Universidad de Pensilvania desarrolló el proyecto Jack, considerado el primer humano virtual. Este software permitía simulaciones interactivas mediante la parametrización del cuerpo humano y su conducta. De este modo:

las máquinas cibernéticas, en especial la máquina de Turing, tenían un nuevo estatus, en la medida en que ya no se trataba de un mecanismo en el sentido cartesiano, pero tampoco de un ser viviente. Se trata más bien de un ser órgano-mecánico: un ser mecánico puesto a funcionar en una forma orgánica. (Hui, 2022, p. 211)

## **La cultura del algoritmo**

Luego de cartografiar la genealogía técnica y conceptual de la IA, resulta necesario preguntarse por su inserción en la vida cotidiana. La IA no solo es un conjunto de artefactos, sino un entorno cultural, que define una *cultura del algoritmo*, que no se limita a procesar información, sino que produce subjetividad, normaliza comportamientos y

reconfigura las relaciones de poder, dentro de un capitalismo informático.

Gran parte de la cultura parece haberse convertido en un asunto de algoritmos. Alexander Galloway (2006) teorizó sobre su capacidad para automatizar entornos socioculturales y producir artefactos visuales como los videojuegos. Más tarde, otros fenómenos culturales se han basado en “programas que nos dimensionan, evalúan lo que queremos, y nos devuelven una experiencia personalizada, al parecer, elaborada solamente a nuestra medida” (Lazer, 2015, p. 1090). Lev Manovich (2013) denomina a esta situación como la cultura del software, que ha colonizado la cotidianidad de millones de personas.

Estas culturas son la antesala de la IA, que representa una convergencia NBIC (nanotecnología, biotecnología, infotecnología y cognotecnología) y ha superado la especulación tecnológica y la ficción, impactando profundamente en aspectos culturales, económicos, tecnológicos, científicos y ambientales. La IA introduce paradigmas organizadores de la vida que apenas comenzamos a vislumbrar, generando temores y glorificaciones que, con frecuencia, opacan la reflexión crítica.

La IA debe entenderse como un proceso capaz de automatizar la percepción humana, lo que la sitúa más allá de las características tradicionales de un autómatas y abre un

debate sobre su naturaleza como algoritmo orientado al reconocimiento de patrones estadísticos. Matteo Pasquinelli y Vladan Joler (2021) denominan este fenómeno como “una era de ciencia ficción estadística sin ningún control político o comunitario”, destacando el carácter autónomo y desregulado de las operaciones algorítmicas contemporáneas.

El aprendizaje maquínico, en sentido estricto, realiza una cartografía basada en algoritmos que procesan información, extendiéndose más allá de los datos perceptibles en internet.

El poder normativo de la IA en el siglo XXI debe ser sometido a escrutinio en tres términos epistémicos: ¿qué significa enmarcar el conocimiento colectivo como patrones, y qué implica trazar espacios vectoriales y distribuciones estadísticas de los comportamientos sociales? (Pasquinelli & Joler, 2021, p. 10)

Esta situación ya había sido descrita por Foucault (2007) al analizar los inicios de la modernidad, donde el poder estadístico era “utilizado para medir las normas sociales y diferenciar el comportamiento normal y anormal” (Foucault, 2007, p. 26). En esa línea, la IA puede entenderse como una nueva forma de *poder de normalización* que se extiende a la burocracia, la educación, la medicina y, por supuesto, a la estadística.

Siguiendo a Foucault, la IA constituye una extensión de las tecnologías disciplinarias. Su capacidad para procesar grandes volúmenes de información configura una forma inédita de vigilancia y control, interpretable como una herramienta sofisticada de lo que Foucault denominó gubernamentalidad: el conjunto de sistemas y técnicas destinados a gobernar la conducta. De este modo, posibilita un nivel de monitoreo e influencia sobre el comportamiento humano muy superior al de los observadores individuales, constituyendo un panóptico no limitado por muros institucionales, sino expandido al ámbito de la interacción digital.

Desde esta perspectiva, la vigilancia propiciada por la IA se concibe como una nueva modalidad de panóptico, donde la observación se descentraliza e internaliza a escala global. La proliferación de estos sistemas erosiona la privacidad y altera las dinámicas entre poder y saber, generando asimetrías de conocimiento explotables por quienes poseen la capacidad de acceder, interpretar y manipular la información. En la visión foucaultiana, al procesar datos personales y contextuales, la IA adquiere la capacidad de predecir, modificar y regular comportamientos en tiempo real. De esta forma, “la norma institucional se ha vuelto una norma computacional: la clasificación de sujetos, cuerpos y comportamientos ya no parece ser un asunto de los registros públicos, sino, por el contrario, de los algoritmos y los centros de datos” (Pasquinelli & Joler, 2021, p. 11).

La esencia de la tecnología no reside en el dispositivo, sino en el proceso creativo y abierto de su uso, que lo convierte en vínculo entre el humano y el mundo. La IA puede entenderse como una nueva forma de *poder de normalización* que, a diferencia de la modernidad clásica, ya no pertenece a los Estados sino a las corporaciones tecnológicas. Anna Schober-de Graaf (2024) y Lotte Philipsen (2024) reflexionan sobre la tendencia a utilizar figuras visuales similares y una iconografía transnacional que homogeneiza las imágenes y las despoja de toda carga crítica. Para Philipsen (2024), la estandarización visual tiene consecuencias éticas, pues reduce la representatividad de minorías y colectivos e impulsa un modo de vida alejado de la pluralidad y diversidad humanas.

Para Bernard Stiegler (2002), comprender las nuevas tecnologías implica ir más allá de una visión meramente instrumental; se trata de la técnica como fusión: no como ente en sí mismo, sino como dos seres independientes pero vinculados de forma inseparable. Stiegler argumenta que nuestra existencia se define por las experiencias y el conocimiento sobre las herramientas que permiten dominar la naturaleza. Hoy, las IA parecen dominar el último vestigio de la inteligencia humana. Ya Martin Heidegger (1997) argumentó que la esencia de la técnica no tenía que ver con lo tecnológico.

Éric Sadin (2020) sostiene que la IA representa el acontecimiento tecnológico más determinante del siglo XXI, no solo por sus capacidades técnicas, sino por su potencia para reorganizar las formas de vida, el pensamiento y las relaciones sociales. Su expansión ha dado lugar a un nuevo régimen de poder cognitivo y moral que pone en crisis los fundamentos del humanismo moderno. Según Eric Sadin (2020) la IA encarna un antihumanismo radical que desplaza gradualmente las capacidades de juicio, decisión y atribución de sentido que tradicionalmente definían al sujeto. Lo que está en juego no es únicamente la automatización del trabajo, sino, de manera más profunda, la automatización del pensamiento. Sadin advierte que la humanidad ha ingresado en una fase dominada por la figura del *homo algorithmicus*, cuya existencia se encuentra regulada por la cuantificación de datos y el rendimiento medido algorítmicamente. La IA opera sobre la base del big data, dando origen a una gobernanza algorítmica que desplaza la subjetividad en favor de una racionalidad estadística y predictiva. Esto convierte al big data y la IA en formas de cuantificar tendencias y patrones que “permiten hacer pronósticos sobre el comportamiento humano. De este modo, el futuro se convierte en predecible y controlable” (Han, 2014, p. 25).

En este contexto, las aplicaciones inteligentes no solo orientan la acción, sino que prescriben modos de vida. Estas tecnologías dejan de ser herramientas técnicas para convertirse en instancias normativas de la conducta social,



instaurando un orden en el que los algoritmos no solo describen la realidad, sino que la modelan y la gobiernan. La IA deja de funcionar como un instrumento tradicional y pasa a operar como amplificación de las capacidades humanas, aunque bajo una lógica que redefine las jerarquías entre sujeto y máquina. Estos cambios implican, para Sadin, una inversión ontológica: las máquinas tienden a dejar de estar subordinadas a las órdenes humanas para constituirse como una autoridad algorítmica capaz de delegar el sentido moral y práctico a los sistemas automáticos.

Esta situación ha sido impulsada por una ideología tecnoliberal, donde los intereses de corporaciones como Google, Amazon o Meta promueven la idea de una sociedad optimizada por datos, concebida como una utopía perfecta. Dicha visión se inscribe en la tradición positivista que sostenía que los problemas sociales, políticos y económicos podían resolverse mediante procedimientos científicos y cuantificables. En este marco, todo se vuelve susceptible de ser medido, calculado y gestionado algorítmicamente. La automatización moral producida por la IA conduce, según Sadin, a una anestesia de la conciencia: dejamos de pensar en las consecuencias de los actos porque confiamos en la supuesta objetividad del sistema, dejamos de ejercer el juicio ético porque delegamos la decisión en el sistema, confiando en su supuesta neutralidad técnica. Este proceso marca la transición hacia una postética tecnológica, donde el juicio

humano se sustituye por la eficiencia de los modelos predictivos.

Este solucionismo tecnológico, como lo denomina Evgeny Morozov (2016), elimina el conflicto, la incertidumbre y la experiencia humana como centro de la toma de decisiones. Bajo esta lógica, la IA impone un orden de transparencia total y control absoluto, incompatible con la imperfección, el error y la libertad: condiciones esenciales de la existencia humana. Los intereses corporativos, según Kate Crawford (2023), generan formas que replican relaciones de poder previas, unidas a estructuras de automatización y gobernanza opaca. Son algoritmos de opresión que, según Umoja Noble (2018), introducen nuevos sesgos y estereotipos que marginan a ciertos grupos definidos por género y raza, como sostiene Nina Mollema (2024). De esta forma, se instalan ciertas automatizaciones y homogeneizaciones que marginan saberes situados en áreas geográficas específicas, hasta convertirse, como sostiene Eden Medina (2013), en una forma de colonialismo de datos.

## **Por una ética de la Inteligencia Artificial**

El poder normalizador de los algoritmos nos sitúa ante una pregunta: ¿cómo responder éticamente a esta nueva forma de gobierno?, ¿qué principios deben guiar su desarrollo? Este subcapítulo explora las posibles respuestas éticas y normativas.

Para Bostrom (2014), la IA superará eventualmente a la inteligencia humana en diversas actividades. Su avance constituye el corolario de la idea persistente de que las máquinas piensen por sí mismas, superando limitaciones humanas: procesando volúmenes ingentes de información, prescindiendo de elementos como los sentimientos, los inhibidores morales, los derechos humanos y, fundamentalmente, el factor tiempo.

Una afirmación recurrente sostiene que la IA constituye un peligro para la humanidad, incluso una amenaza existencial. Algunos especulan con la posibilidad de que se desarrolle de manera espontánea y descontrolada, comparable a “un cáncer exponencialmente inteligente. Podríamos empezar con algo simple, pero la inteligencia mejora escapando a nuestro control. Antes de que nos demos cuenta, toda la raza humana está luchando por su supervivencia” (Bentley, 2018, p. 8).

En 2002, Bostrom introdujo el concepto de riesgo existencial para argumentar que los resultados adversos de la tecnología podrían desencadenar dos escenarios apocalípticos: la aniquilación o la reducción drástica de la vida inteligente en el planeta. Esta perspectiva resulta relevante para describir los posibles efectos de la IA.

No obstante, para los precursores de la IA, la clave no reside en la cantidad de recursos, sino en la capacidad de

organizarlos eficazmente. Como señalan Aleena Garner y Mark Mayford (2012), el verdadero desafío consiste en organizar estos recursos mediante algoritmos óptimos, no en acumular neuronas o chips.

Una lectura crítica y premonitoria de estos debates se encuentra en la novela *Yo, robot* (1950) de Isaac Asimov, que trascendió los ámbitos políticos, filosófico y moral al anticipar los problemas que enfrentaría la humanidad si los robots alcanzaran inteligencia y autonomía. Asimismo, vislumbró el poder de las corporaciones privadas al monopolizar la investigación en este campo. Asimov estableció las tres leyes de la robótica, que permanecen vigentes en los debates éticos:

- Un robot no debe dañar a un ser humano o, por su inacción, permitir que un ser humano sufra daño.
- Un robot debe obedecer las órdenes de los seres humanos, excepto cuando estas órdenes contravengan la Primera Ley.
- Un robot debe proteger su propia existencia, siempre que esta protección no entre en conflicto con la Primera o Segunda Ley.

El ascenso de la IA plantea diversos dilemas, siendo los más apremiantes la posibilidad de que estas inteligencias desarrollen un lenguaje independiente de sus creadores y la capacidad de autorreplicarse. Estas preocupaciones fueron

anticipadas en películas como *2001: Una odisea del espacio* (1968), *Terminator* (1984) o *Her* (2013).

La velocidad de los ordenadores se duplica en periodos muy cortos. Según los criterios de la IA, a mayor velocidad, mayor capacidad para crear una inteligencia que supere a la humana. Sin embargo, dicha velocidad se ve frenada por los debates éticos sobre el impacto exponencial de su uso, por lo que toda experimentación tecnológica se basa en la prueba constante.

La existencia de la codificación humana y de datos, junto con el aumento de la velocidad, la capacidad de almacenamiento y el procesamiento de datos, ha hecho posible crear entornos digitales capaces de adaptarse y responder a situaciones humanas, hasta proyectar nuestra corporalidad e imagen en mundos virtuales. A esto se añade la creación de prótesis electromecánicas y la manipulación genética. Este nuevo entorno permite poner en funcionamiento sistemas autónomos capaces de conocer los entornos donde se ubican.

Shannon Vallor (2024) sostiene que los sistemas de IA funcionan como espejos de nuestra inteligencia: ejecutan las órdenes impartidas y sus resultados dependen de la formulación de los prompts, los cuales deben construirse a partir de diversas lógicas, entre ellas la instrucción de las tareas y la contextualización de la búsqueda. Estas

observaciones demuestran que la IA no piensa, sino que obedece órdenes, operando sin comprensión, conciencia, emociones ni sensibilidad.

Para Sadin (2020), la pandemia de COVID-19 nos enfrentó a una doble situación: evidenció nuestra vulnerabilidad como especie biológica y aceleró la transferencia de nuestra vida al mundo digital, una tendencia ya prevista pero intensificada por la crisis sanitaria. Este proceso permitió que nuestra existencia se desarrollara en lo que el autor denomina sociedad fantasma: un mundo espectral, impersonal y desconectado.

En el ámbito tecnológico, Mark Zuckerberg plantea la idea de un metaverso digital, concepto originado en la ciencia ficción que hoy parece convertirse en realidad, anticipado por el universo distópico de *Matrix* (1999). El metaverso de Zuckerberg seduce al prometer el cumplimiento de nuestros anhelos sin coerción. Sadin (2020) interpreta esto como el inicio de una era de integrismo digital regida por algoritmos y azar, donde toda experiencia se reduce a datos cuantificables.

Desde la perspectiva de Baudrillard (1978), asistimos a un mundo de simulacros donde imágenes y símbolos han abandonado su función representativa para constituir una hiperrealidad: una construcción artificial que no solo distorsiona lo real, sino que lo reemplaza por completo.

Luciano Floridi (2024) ofrece una interpretación filosófica de la IA concebida como una tecnología central del mundo contemporáneo, que debe entenderse no solo como “una nueva forma de inteligencia, sino como una nueva forma de agencia” (Floridi, 2024, p. 137). De este modo, la IA se ubica como una tecnología de tercer orden: no solo permite acceder a la información, como la imprenta (primer orden) o la computadora (segundo orden), sino que posee la capacidad de interactuar y generar un nuevo ecosistema digital que “transforma radicalmente la realidad porque crea nuevos entornos que habitamos y nuevas formas de agencia con las que interactuamos” (Floridi, 2024, p. 57). Esta tecnología tiene la capacidad de acoplarse, desacoplarse o reacoplarse, lo que conduce a una “reontologización y reepistemologización de nuestra realidad”.

La tesis de Floridi sitúa a la IA como un triunfo de la agencia: desacopla la resolución de problemas de la inteligencia biológica y permite ejecutar tareas de manera efectiva. Así, la IA promueve la búsqueda de resultados sin reproducir inteligencia biológica, operando como un juego algorítmico de información. Su rapidez para alcanzar objetivos ha contribuido a su popularidad, conquistando un espacio significativo en tareas y problemas antes exclusivamente humanos. El futuro de la IA reside en “ludificar y envolver: es cuestión de diseñar, o a veces rediseñar, las realidades con las que tratamos” (Floridi, 2024,

p. 130). En cuanto a la ética, propone cinco principios fundamentales para la IA, heredados de la bioética:

- Beneficencia: promover el bienestar y preservar la dignidad.
- No maleficencia: proteger la privacidad y seguridad de los datos de los usuarios.
- Autonomía: garantizar el liderazgo humano.
- Justicia: fomentar la solidaridad y prevenir la injusticia.
- Explicabilidad: asegurar la inteligibilidad de las tecnologías inteligentes.

Desde un enfoque psicoanalítico y crítico-ideológico, Slavoj Žižek (2021) examina los posibles riesgos éticos y humanistas de la inteligencia artificial, entre los que identifica:

- El desencantamiento de la realidad humana;
- Las amenazas a la autonomía individual;
- La deshumanización y pérdida de autenticidad; y
- El peligro de la complacencia tecnológica.



De esta forma:

la IA representa una amenaza fundamental para la experiencia humana tal como la conocemos. Al algoritmizar y digitalizar cada vez más dimensiones de nuestra realidad social, se corre el riesgo de erosionar nuestro sentido de nosotros mismos como agentes creativos y espontáneos. (Žižek, 2021, p. 33)

El éxito de la IA podría interpretarse como un efecto del desgaste de la vida cotidiana: un “desencantamiento” del mundo físico que se remonta a la Ilustración y que ha ido erosionando la capacidad humana de imaginación, creatividad y espontaneidad.

Lo preocupante no es solo el control directo sobre nuestras mentes y cuerpos, sino más bien el escenario indirecto en el que nuestra experiencia social es sutilmente manipulada por algoritmos secretos que orquestan nuestros encuentros, relaciones e ideas. De esta manera, la realidad vivida se vuelve cada vez más estandarizada y aplanada, perdiendo su profundidad espontánea. (Žižek, 2021, p. 42)

La promesa de la IA, obtener beneficios mediante la predicción de patrones, la personalización de contenidos y la automatización de tareas, socava, según Žižek, la noción misma de libertad humana, calificándola incluso como una

“dictadura amable”: un sistema que seduce a los sujetos con la promesa de conocimiento y eficiencia, pero que en realidad responde a intereses corporativos y tecnológicos.

Otro de los problemas asociados al uso de la IA es el de la “extracción”, concepto que, según Kate Crawford (2023), presenta dos modos de acción: 1) la abstracción de las condiciones materiales, y 2) la acumulación de información y recursos. Los efectos derivados del uso de la IA generan un nuevo tipo de impacto social que se manifiesta en tres esferas principales: los beneficiarios del desarrollo tecnológico, los usuarios que encuentran en la IA un medio de acceso rápido y sencillo al conocimiento, y los desposeídos que permanecen en una posición de desventaja. De este modo, se revela su cara oculta:

El ciclo de vida de un sistema de IA tiene muchas cadenas fractales: formas de explotación de mano de obra y de recursos naturales, y concentraciones masivas de poder corporativo y geopolítico. Y a lo largo de toda la cadena, un continuo consumo en amplia escala de energía que mantiene el ciclo funcionando. (Crawford, 2023, p. 61)

Así, se configura un riesgo ético: nuestras vidas podrían verse automatizadas por el uso de la IA, lo que equivaldría a vivir bajo una lógica del cálculo desprovista de virtudes intelectuales y morales: “lo que la IA pone en mayor

peligro no es la mera supervivencia de nuestra especie, sino nuestra propia autocomprensión de la perpetua libertad y responsabilidad” (Vallor, 2024, p. 195).

El dilema ético de la IA se articula en torno a la pregunta de qué ocurre cuando la tecnología comienza a moldear nuestras vidas de formas inesperadas. Para Stuart Russell (2019), existe un profundo temor respecto al uso militar de los sistemas de IA, en especial en la creación de sistemas de armas autónomas letales (LAWS, por sus siglas en inglés), definidos como dispositivos capaces de operar sin control humano. Russell advierte que estas podrían violar principios fundamentales de la dignidad humana al permitir que las máquinas decidan a quién matar.

Por su parte, Russ Altman y Jun Kim (2025) sostienen que la IA posee un enorme potencial para acelerar el descubrimiento científico, particularmente en los campos de la biología y la medicina. Los sistemas de IA facilitan la interpretación de grandes volúmenes de datos biológicos; sin embargo, este avance suscita preocupaciones sobre la posibilidad de un nuevo apartheid tecnológico que podría exacerbar las desigualdades sanitarias a escala global.

Manuela Veloso (2023) plantea que los seres humanos integran de manera orgánica la percepción, la cognición y la acción, y que la IA puede reforzar este proceso al permitir evaluar el estado del mundo en tiempo real. En esta

integración creciente, nuestros cerebros comienzan a delegar parte de la toma de decisiones a los algoritmos, lo que redefine el modo en que pensamos y actuamos.

## **El aceleracionismo: hacia una teoría de la fluidez radical de la IA**

Frente a los dilemas éticos propuestos las diversas respuestas tienden a oscilar entre el optimismo tecnológico y el llamamiento a la regulación. No obstante, hay una corriente de pensamiento que rechaza estas dos posiciones y propone una vía radical: el aceleracionismo que constituye un horizonte ideológico que subyace a promesas y temores de la IA, y que, bajo sus dos vertientes, la de derechas y la de izquierdas, condensa las preguntas sobre nuestro futuro tecnológico.

La filosofía del aceleracionismo, surgida en Inglaterra durante la década de 1990, concretamente en 1995, en la Universidad de Warwick, cuando se fundó la Cybernetic Culture Research Unit (CCRU) en el Departamento de Filosofía, bajo la dirección de Sadie Plant y Nick Land. Este pensamiento se caracteriza por su capacidad de fragmentarse mientras conserva un núcleo filosófico que reinterpreta a Marx, Deleuze y Guattari. Para Armen Avanessian y Mauro Reis (2017), el aceleracionismo constituye una respuesta política radical al capitalismo tecnológico que trasciende la

mera protesta para buscar la aceleración de sus procesos contradictorios hasta provocar su autodestrucción.

El aceleracionismo analiza los avances tecnológicos dentro del capitalismo tardío y plantea interrogantes cruciales sobre el presente y futuro de la humanidad: ¿puede la tecnología convertirse en una palanca para la emancipación humana? ¿Qué roles desempeñan el capitalismo, la tecnología y los seres humanos en la definición de las condiciones sociales futuras? La fragmentación del movimiento surge de las diversas respuestas a estas preguntas, que abarcan desde el fatalismo tecnológico hasta la utopía romántica de la salvación tecnológica.

Bajo estos argumentos, el aceleracionismo parte de la premisa de que no es posible volver al pasado ni aislarse del mundo. Desde una posición marxista, no puede escaparse del capitalismo; debe superarse mediante la aceleración de los procesos políticos, económicos y tecnológicos, sin eludir las cuestiones sobre la capacidad humana para decidir dichos procesos, sus consecuencias y sus posibles víctimas.

El aceleracionismo retoma de Marx la concepción del capitalismo como una contradicción que genera crisis cada vez más peligrosas, situando a los sujetos en condiciones de vulnerabilidad insostenible. Para escapar de esta situación, los aceleracionistas proponen: 1) adelantar la revolución mediante el despertar del proletariado para evitar

sufrimientos innecesarios, y 2) utilizar el capitalismo como arma contra sí mismo. En sociedades donde la mayoría está seducida por el capitalismo generando una ceguera sobre sus contradicciones, como advirtieron Adorno y Horkheimer (1998), los aceleracionistas proponen no regularlo, sino impulsarlo hasta su límite para que se autodestruya. Los aportes de Deleuze y Guattari en *El Anti-Edipo* (2004) defienden la alternativa revolucionaria de acelerar los procesos del capitalismo y la tecnología que aún no han completado la territorialización total del planeta.

Para Nick Land (2021), figura clave del aceleracionismo de derecha, el capitalismo es una inteligencia no humana proveniente del futuro, por lo que no debe contenerse, sino liberarse de cualquier regulación que limite su desarrollo. La tecnología cobra importancia crucial al crear formas de inteligencia superior que trasciendan lo humano, la llamada singularidad tecnológica. El capitalismo es una fuerza imparable que se autoorganiza y autorregula, donde la emancipación no es humana, sino la de los medios de producción.

Inspirándose en Nietzsche, Deleuze y Guattari, Land reflexiona sobre la filosofía moderna como herramienta de represión que separa el deseo de la acción, cercena la imaginación natural y somete al animal humano a la ley universal, sofocando la vida. Desde una lectura de Deleuze y Guattari, apuesta por una filosofía materialista y atea que

rechaza la trascendencia: no hay dialéctica ni trascendencia, solo máquinas y flujos.

El pensamiento de Land se encuadra dentro del neorreaccionarismo, movimiento de extrema derecha y antidemocrático que se presenta como antítesis de la Ilustración y aboga por abandonar la sociedad postcapitalista para centrarse en un futuro dominado por la tecnología y la inteligencia artificial.

El Manifiesto por una política aceleracionista de Alex Williams y Nick Srnicek (2013) inició un amplio debate filosófico en torno al llamado *aceleracionismo de izquierdas* y el *aceleracionismo de derechas*. La diferencia radica en que la vertiente de derechas se asienta sobre una concepción metafísica de la historia orientada hacia su propio colapso, vinculado paradójicamente con el progreso; mientras que la de izquierdas adopta un enfoque político que busca intervenir activamente para subvertir las relaciones sociales y conducirlas hacia una era poscapitalista. “Frente al relato del fin de la historia y la inexistencia de alternativas proclamadas por el neoliberalismo, el aceleracionismo se reivindica como una estrategia política orientada hacia el futuro” (Acosta, 2024, p. 263). Desde la izquierda, el aceleracionismo se entiende como una “estrategia política” que se diferencia de la izquierda liberal, en la medida en que sostiene que la dominación instrumental del capitalismo contiene un potencial emancipador hacia un futuro postindustrial.

A partir de entonces, las distintas vías de transformación social se han centrado en formas de resistencia ante la acelerada expansión del neoliberalismo y del capitalismo financiero, lo que sus críticos denominan con cierto desdén 'políticas folks', o en un lenguaje más común, simples políticas keynesianas. (Acosta, 2024, p. 264)

Mark Fisher (2016), aceleracionista de izquierda, sostiene que el capitalismo no es una fuerza desterritorializada, pues posee dinámicas que lo limitan; el capitalismo no tiene velocidad de aceleración, sino un ralentizador que cancela el futuro. Argumenta que usar la tecnología no significa perpetuar el capitalismo, sino exponer sus contradicciones para construir una sociedad más justa, lo que implica mirar al futuro sin idealizar el pasado. Fisher recurre a la frase de Fredric Jameson es más fácil imaginar el fin del mundo que el fin del capitalismo para describir el realismo capitalista, concepto que adapta de una corriente artística alemana surgida en Berlín en 1963 para desafiar la dominación del arte pop estadounidense.

Esta corriente combinaba las ideologías del arte pop y el realismo socialista para criticar la cultura de consumo y el poder político capitalista. Fisher emplea este concepto para describir una dimensión ontológica del capitalismo tardío, añadiendo una posición política y económica de nuestro



tiempo que entiende los problemas públicos desde intereses privados regulados por mecanismos de mercado.

La globalización, el desplazamiento de las manufacturas por la computarización, la precarización del trabajo y la intensificación de la cultura de consumo. Estos rasgos constituyen el fundamento invisible de la realidad incontrovertible y ostensiblemente pospolítica sobre la que descansa el realismo capitalista. (Fisher, 2016, p. 129)

Para Fisher, asistimos a una desaceleración cultural iniciada en la década de 1980, marcada por la inercia, repetición y retrospección culturales, y acentuada con la caída del Muro de Berlín. Este contexto creó escenarios propicios para que las biotecnologías penetraran en lo social ofreciendo sus bondades para mejorar la vida. La desaceleración cultural presentó un mundo gris donde las relaciones sociales en crisis necesitaban nuevas motivaciones, que la tecnología, especialmente la mediática e informática, intentó cubrir. “Quiero situar el aceleracionismo no como una forma herética del marxismo, sino como un intento de converger con, intensificar y politizar las dimensiones más desafiantes y experimentales de la cultura popular” (Fisher, 2017, p. 158).

Para muchos de los autores aquí abordados, el crecimiento tecnológico parece haberse vuelto incontrolable e irreversible, lo que sitúa a la civilización ante un punto de no

retorno cuyos resultados aún resultan imprevisibles. Hemos llegado a un punto de no retorno cuyos resultados aún resultan imprevisibles. La inteligencia artificial, que en sus inicios pertenecía al ámbito de la ciencia ficción, se ha convertido hoy en una presencia cotidiana.

Ray Kurzweil en *The Singularity Is Near* (2005) combina lo especulativo con lo científico hasta predecir que las máquinas desarrollarán capacidades lingüísticas equivalentes a las humanas hacia 2029 y que, para 2045, alcanzarán una inteligencia colectiva muy superior a la del ser humano. Para Kurzweil, la singularidad está guiada por las fuerzas del mercado y representa una visión liberal y optimista, fundada en la idea de que el capitalismo y el crecimiento tecnológico forman parte del proceso de emancipación humana que trasciende los límites de la biología. En cambio, Land (2021) considera que el capitalismo mismo es una forma de inteligencia artificial: los mercados funcionan como sistemas proto-IA que se autoorganizan, evolucionan y buscan la eficiencia de manera autónoma.

Por su parte, Bostrom adopta una perspectiva centrada en el riesgo existencial y en los posibles desenlaces a largo plazo del advenimiento de una superinteligencia artificial general. La lógica interna del capitalismo concibe la automatización y la maximización de la eficiencia como necesidades inherentes al desarrollo de la IA, orientadas a

reemplazar la fuerza de trabajo en todas las esferas de la actividad humana, lo que podría conducir a un colapso social de proporciones impredecibles.

Algunas instituciones internacionales formulan recomendaciones sobre principios mínimos de justicia, responsabilidad e igualdad que rijan los próximos pasos en la implementación y adopción social de la IA. La Unión Europea ha comenzado a implementar normativas basadas en la transparencia y la explicabilidad para conseguir una gobernanza ética.

Los sistemas de IA plantean nuevos tipos de cuestiones éticas que incluyen, aunque no exclusivamente, su impacto en la adopción de decisiones, el empleo y el trabajo, la interacción social, la atención de la salud, la educación, los medios de comunicación, el acceso a la información, la brecha digital, la protección del consumidor y de los datos personales, el medio ambiente, la democracia, el estado de derecho, la seguridad y el mantenimiento del orden, el doble uso y los derechos humanos y las libertades fundamentales, incluidas la libertad de expresión, la privacidad y la no discriminación. (UNESCO, 2022, p. 10)

Están más que servidas las implicaciones de esta competición entre la inteligencia humana y la artificial para generar nuevas posibilidades de vida. En cualquier caso,

deberían producirse bajo criterios claros que hagan la inteligencia artificial éticamente sostenible (Terrones Rodríguez, 2022), responsable (Jonas, 1994, 2001) y atenta al cuidado de todos los grupos, colectivos y pueblos indígenas (Gilligan, 1985, 1993).

## **Conclusión**

La irrupción de la IA señala un cambio fundamental en la forma de vida que desafía la visión antropocéntrica dominante durante siglos. La creencia de que los humanos somos el centro del universo intelectual y moral pierde referencia cada día, permitiendo a la IA convertirse en un espejo. Incluso se ha llegado a argumentar, como hace Flavia Acosta (2024) al hilo de Martins, que el Antropoceno ha conducido al fin del antropocentrismo, allanando el camino hacia lo que podríamos denominar, con todas las precauciones, un incipiente *codocentrismo*: la hipótesis de que el algoritmo tiende a convertirse en un nuevo centro simbólico de la vida. La injerencia de la IA se extiende más allá del tejido social. Por ello, urge un debate público transparente, enmarcado como una necesidad para la supervivencia humana. Todo esto nos ha conducido a una encrucijada filosófica, moral y práctica. Las cuestiones éticas ya no se limitan a las interacciones entre humanos; ahora debemos navegar en un mundo donde nuestros principales interlocutores y adversarios pueden ser inteligencias no humanas de nuestra creación.

La nueva realidad exige reconocer que la IA desafía los asuntos militares, la gestión ambiental y la gobernanza social. Requiere un debate público transparente y continuo. El fin no puede ser dominar esta nueva inteligencia, sino aprender a coexistir con ella. Si no podemos afrontar esta situación, será por nuestra incapacidad para evolucionar nuestras acciones y pensamientos.

## Referencias

- Acosta, L. (2024). Aceleracionismo y marxismo: Un debate en torno a la supervivencia del trabajo en la era de la inteligencia artificial. En J. C. Ruíz Sánchez & M. Bermúdez Vázquez (Eds.), *Territorios del pensamiento* (pp. 263–280). Dykinson.
- Adorno, T., & Horkheimer, M. (1998). *Dialéctica de la Ilustración*. Trotta.
- Altman, R., & Kim, J. (2025). Semi-supervised data-integrated feature importance enhances performance and interpretability of biological classification tasks. *Bioinformatics*, 41(Suppl. 1), i373–i381. <https://doi.org/10.1093/bioinformatics/btaf190>
- Avanessian, A., & Reis, M. (Eds.). (2017). *Aceleracionismo: Estrategias para una transición hacia el postcapitalismo*. Caja Negra.
- Baudrillard, J. (1978). *Cultura y simulacro*. Kairós.
- Bengio, Y. (2023). AI and catastrophic risk. *Journal of Democracy*, 34(4), 5–16.

<https://www.journalofdemocracy.org/ai-and-catastrophic-risk/>

- Bentley, P. (2018). *¿Debemos temer a la inteligencia artificial?* Dirección General de Servicios de Estudios Parlamentarios, Parlamento Europeo.
- Bostrom, N. (2011). Argumentos de razón técnica: Sociedad, y filosofía de la tecnología. *Revista Española de Ciencia, Tecnología, 14,* 157–191.  
[https://institucional.us.es/revistas/argumentos/14/art\\_7.pdf](https://institucional.us.es/revistas/argumentos/14/art_7.pdf)
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Braidotti, R. (2020). *El conocimiento posthumano*. Gedisa.
- Crawford, K. (2023). *Atlas de la inteligencia artificial: Poder, política y costos planetarios*. Fondo de Cultura Económica.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray.
- Deleuze, G., & Guattari, F. (2004). *El Anti-Edipo: Capitalismo y esquizofrenia*. Paidós.
- Dreyfus, H. L. (1965). *Alchemy and artificial intelligence* (Paper P-3244). RAND Corporation.
- Echeverría, J. (2003). *La revolución tecnocientífica*. Fondo de Cultura Económica.
- Fisher, M. (2016). *Realismo capitalista*. Caja Negra.

- Fisher, M. (2017). Una revolución social y psíquica de magnitud casi inconcebible: Los interrumpidos sueños aceleracionistas de la cultura popular. En A. Avanessian & M. Reis (Eds.), *Aceleracionismo: Estrategias para una transición hacia el postcapitalismo* (pp. 155–174). Caja Negra.
- Floridi, L. (2024). *Ética de la inteligencia artificial*. Herder.
- Foucault, M. (2007). *Los anormales*. Fondo de Cultura Económica.
- Galloway, A. R. (2006). *Gaming: Essays on algorithmic culture*. University of Minnesota Press.
- Garner, A. R., & Mayford, M. (2012). Generation of a synthetic memory trace. *Science*, 335(6075), 1513–1516. <https://doi.org/10.1126/science.1214985>
- Gilligan, C. (1985). *La moral y la teoría*. Fondo de Cultura Económica.
- Gilligan, C. (1993). Reply to critics. En M. J. Larrabee (Ed.), *An ethic of care: Feminist and interdisciplinary perspectives* (pp. 207–214). Routledge.
- Halpin, H. (2025). Artificial intelligence versus collective intelligence. *AI & Society*, 40(9), 4589–4604. <https://doi.org/10.1007/s00146-025-02240-x>
- Han, B.-C. (2014). *Psicopolítica*. Herder.
- Heidegger, M. (1997). La pregunta por la técnica. En *Filosofía, ciencia y técnica*. Editorial Universitaria.
- Hui, Y. (2022). *Recursividad y contingencia*. Caja Negra.

- Jonas, H. (1994). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica*. Círculo de Lectores.
- Jonas, H. (2001). *Más cerca del perverso fin y otros diálogos y ensayos*. Los Libros de la Catarata.
- Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Viking Press.
- Land, N. (2014). *The dark enlightenment*. Recuperado el 15 de mayo de 2025, de <http://www.thedarkenlightenment.com/the-dark-enlightenment-by-nick-land/>
- Land, N. (2021). *Teleoplexia: Ensayos sobre aceleracionismo y horror*. Holobionte.
- Lazer, D. (2015). The rise of the social algorithm. *Science*, 348(6239), 1090–1091. <https://doi.org/10.1126/science.aab1422>
- Manovich, L. (2013). *Software takes command: Extending the language of new media*. Bloomsbury Academic.
- Martins, H. (2018). *The technocene: Reflections on bodies and markets*. Anthem Press.
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (2006). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, 27(4), 12–14. <https://doi.org/10.1609/aimag.v27i4.1904>



- McLuhan, M. (1996). *Comprender los medios de comunicación: Las extensiones del ser humano*. Paidós.
- Medina, E. (2013). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and resistant imaginations*. Oxford University Press.
- Mollema, W. (2024). Decolonial AI as disenclosure. *Open Journal of Social Sciences*, 12, 574–603. <https://doi.org/10.4236/jss.2024.122032>
- Morozov, E. (2016). *La locura del solucionismo tecnológico*. Katz Editores.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Pasquinelli, M., & Joler, V. (2021). El Nooscopio de manifiesto. *laFuga*, 25. <https://lafuga.cl/el-nooscopio-de-manifiesto/1053>
- Philipsen, L. (2024). 'Natural image processing': Aesthetic ideologies of text-to-image AI. En *Aesthetics of Digital Image Syntheses*. Universität Klagenfurt. <https://arthist.net/archive/42851>
- Rifkin, J. (2011). *La tercera revolución industrial*. Paidós.
- Russell, S. (2019). *Human compatibility: Artificial intelligence and the problem of control*. Viking Press.
- Sadin, É. (2020). *La inteligencia artificial o el desafío del siglo: Anatomía de un antihumanismo radical*. Caja Negra.

- Schober, A. (2024). I-figures in concrete scenes: Patterns of audience address and their realisation in contemporary art. En *Aesthetics of Digital Image Syntheses*. Universität Klagenfurt.  
<https://arthist.net/archive/42851>
- Srnicek, N., & Williams, A. (2015). *Inventing the future: Postcapitalism and a world without work*. Verso.
- Stiegler, B. (2002). *La técnica y el tiempo*. Hiru.
- Terrones Rodríguez, A. L. (2022). Ética para la inteligencia artificial sostenible. *Arbor*, 198(806), a683.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.  
<https://www.hec.edu/sites/default/files/documents/Computing%20Machinery%20and%20Intelligence%20oby%20Alan%20Turing.pdf>
- UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial*.  
[https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa)
- Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in an age of machine thinking*. Oxford University Press.
- Veloso, M. (2023). *Autonomous AI agents in education: A cognitive approach*. Springer.
- Williams, A., & Srnicek, N. (2013, 14 de mayo). #Accelerate: Manifesto for an accelerationist politics. *Critical Legal Thinking*.

<https://criticallegalthinking.com/2013/05/14/accelerate-manifesto-for-an-accelerationist-politics/>

Žižek, S. (2021). *Como un ladrón en pleno día: El poder en la era de la poshumanidad*. Anagrama.

# Gubernamentalidad algorítmica: líneas de fuga para una crítica del poder digital contemporáneo<sup>42</sup>

Algorithmic Governmentality: Lines of Flight for a  
Critique of Contemporary Digital Power

Ivonne Jiménez Jiménez<sup>43</sup>

*Escribir con slogans: ¡Haced rizoma y no raíz, no  
plantéis nunca! ¡No sembréis, horadad! ¡No seáis  
ni uno ni múltiple, sed multiplicidades! ¡Haced la  
línea, no el punto! La velocidad transforma el*

---

<sup>42</sup> El presente capítulo surge como producto de la reflexión teórica de la tesis doctoral en Filosofía que adelanta la autora, en la cual la “gubernamentalidad algorítmica” opera como una de las categorías centrales de la investigación.

<sup>43</sup> Universidad Santo Tomás. <https://orcid.org/0000-0003-4070-5294>

*punto en línea. ¡Sed rápidos, incluso sin moveros!*  
*Línea de suerte, línea de cadera, línea de fuga*

— Deleuze y Guattari (1980)

## **Resumen**

Este capítulo examina la transformación del poder contemporáneo hacia una gubernamentalidad algorítmica, un régimen predictivo que administra la vida mediante el seguimiento continuo de datos. Para abordar críticamente este fenómeno, se recurre al concepto filosófico de líneas de fuga (Deleuze y Guattari) —entendidas operativamente como trayectorias tensionales y vectores de ruptura, alejándose de una simple categorización temática—. Mediante una revisión bibliográfica exhaustiva (2014-2025), el texto supera la dispersión de un inventario descriptivo para proponer una cartografía analítica articulada en cuatro grandes líneas estructurantes: 1) subjetivación biopolítica, 2) vigilancia y control, 3) opacidad y manipulación, y 4) desigualdad estructural y resistencias. A estos ejes teóricos se subordinan dominios empíricos abordados a través de casos que funcionan estrictamente como viñetas ejemplificadoras. El artículo demuestra cómo la maquinaria algorítmica captura y “reestratifica” incesantemente las fisuras del sistema, ofreciendo una lectura filosófica rigurosa sobre los efectos del poder digital y los horizontes de emancipación en la actualidad.

**Palabras clave:** Gubernamentalidad algorítmica; líneas de fuga; poder digital; vigilancia algorítmica; opacidad; desigualdad.

## **Abstract**

This chapter examines the transformation of contemporary power toward an algorithmic governmentality, a predictive regime that manages life through continuous data tracking. To critically address this phenomenon, the philosophical concept of lines of flight (Deleuze and Guattari) is employed—understood operationally as tensional trajectories and vectors of rupture, moving away from simple thematic categorization. Through an exhaustive literature review (2014–2025), the text overcomes the dispersion of a descriptive inventory to propose an analytical cartography articulated around four major structuring lines: 1) biopolitical subjectivation, 2) surveillance and control, 3) opacity and manipulation, and 4) structural inequality and resistances. Empirical domains are subordinated to these theoretical axes and addressed through cases that strictly serve as exemplifying vignettes. The article demonstrates how the algorithmic machinery incessantly captures and “re-stratifies” the fissures of the system, offering a rigorous philosophical reading on the effects of digital power and the horizons of emancipation today.

**Keywords:** Algorithmic governmentality; lines of flight; digital power; algorithmic surveillance; subjectivity.

## **Introducción**

En las últimas décadas, la sociedad de la información ha transitado hacia un régimen de poder que ya no se sustenta en el encierro disciplinario ni en la prerrogativa soberana, sino en la digitalización de la vida y en el seguimiento continuo de datos. Este nuevo poder, al que denominamos gubernamentalidad algorítmica (en adelante GA), no ordena ni prohíbe de forma explícita: anticipa conductas, administra posibilidades y orienta acciones mediante el cálculo estadístico de lo probable. La lógica institucional cede ante un control inmanente, inscrito en redes, plataformas y dispositivos que penetran la cotidianidad y transforman la subjetividad en perfiles técnicos, modulados por algoritmos que clasifican, predicen y normalizan. Uno de los giros decisivos de esta transformación es el paso del territorio al perfil: ya no se gobiernan cuerpos en espacios delimitados, sino flujos de datos que convierten la identidad en una interfaz performativa. Este desplazamiento implica una reorganización de la subjetividad, donde deseos, emociones y decisiones se tornan objetos de gestión digital. La GA encarna así un gobierno sin sujeto visible, un poder predictivo que se adhiere a la vida y la regula desde dentro.

Para abordar críticamente este fenómeno, se recurre al concepto de líneas de fuga (en adelante LF) acuñado por Deleuze y Guattari (2010) como herramienta analítica capitular. En su cartografía del poder, los autores explican que, frente a las estructuras de control establecidas, “se produce una ruptura, se traza una línea de fuga, pero siempre existe el riesgo de que reaparezcan en ella organizaciones que reestratifican el conjunto” (p. 15). Aterrizando esta premisa filosófica a la época digital, en el contexto de esta escritura las LF no se entienden como simples áreas de impacto, vías de escape románticas o categorías temáticas. Por el contrario, se definen operativamente como trayectorias tensionales y vectores de ruptura. Estas líneas permiten mapear la doble dinámica de la GA: por un lado, los modos en que la vida y las resistencias intentan fisurar el sistema; y por otro, la capacidad de la racionalidad algorítmica para “reestratificar” esos escapes, capturándolos incesantemente para convertirlos en nuevas métricas de control.

La metodología empleada consistió en una revisión bibliográfica crítica de más de cincuenta fuentes especializadas (2014–2025), seguida de un proceso de codificación analítica orientada a evitar la dispersión de un mero inventario descriptivo. Esto permitió identificar y sistematizar la maquinaria algorítmica en cuatro grandes líneas de fuga estructurantes: la subjetivación biopolítica, la vigilancia y el control, la opacidad y manipulación, y la desigualdad estructural. A estas líneas de tensión teóricas se



subordinan dominios empíricos específicos (como la economía de plataformas, la educación o el periodismo), abordados en este texto a través de casos que fungen estrictamente como viñetas ejemplificadoras de las rupturas descritas. La sistematización articula así dimensiones teóricas y empíricas para construir una cartografía crítica de la GA, con el objetivo de ofrecer una lectura filosófica rigurosa sobre sus efectos en la vida contemporánea y los horizontes que abre para pensar el presente.

### **Racionalidad algorítmica: del perfil al gobierno digital**

Entender la GA como forma de pensar y ejercer el poder exige revisar sus mecanismos y diferencias respecto a los modelos clásicos. Como advierte Rodríguez (2019) “transcurre en la vigilancia distribuida e inmanente a través de los perfiles de redes” (p. 359). Si el poder disciplinario se fundaba en instituciones cerradas y panópticas, la lógica algorítmica se despliega en redes digitales, operando de manera ubicua, automatizada y continua. Esta mutación se aproxima a la noción de Rouvroy (2020), para quien la GA es “el procesamiento algorítmico de grandes conjuntos de datos en lugar de en la política, la ley y las normas sociales” (p. 1). Se configura así un poder que sustituye la norma por la expectativa y gobierna optimizando conductas sin coerción visible. Según Rouvroy y Berns (2013), GA designa “un cierto tipo de racionalidad (a) normativa o (a) política basada en la

recolección, agregación y análisis automatizado de datos en cantidad masiva, con el fin de modelizar, anticipar y afectar por adelantado los comportamientos posibles” (p. 173). El poder se infiltra en los flujos de lo posible, transformándose en técnico, predictivo y preventivo.

Amoore (2020) subraya que los algoritmos son más que administradores de datos: actúan como instrumentos de percepción y decisión, clasificando riesgos y poblaciones en tiempo real. La GA desplaza el gobierno basado en estadísticas hacia un control enfocado en lo nuevo, lo excepcional y lo posible, trabajando con datos en movimiento. Rodríguez (2019) caracteriza este viraje como el paso del territorio al perfil digital, donde ya no se vigilan cuerpos, sino patrones de datos que permiten actuar “sobre similares” (p. 357). El perfil se convierte en objeto político central, lugar donde se inscribe la normatividad y se modulan conductas. Los datos no representan lo que es, sino lo que podría ser. Como indica el propio autor, los perfiles producen “un efecto de identidad, en un sentido puntual y provisorio, una vez que no atienden a criterios de verdad y falsedad, sino de performatividad” (Rodríguez, 2019, p. 357). Así, la subjetividad se torna interfaz performativa: ya no habla el sujeto, sino sus datos.

En este régimen, la vigilancia se distribuye entre plataformas, sensores y sistemas de recomendación. El individuo mismo “lleva consigo aquello que lo vigila y a través

de lo cual puede vigilar” (Rodríguez, 2019, p. 358). La subjetividad se representa a través de huellas digitales, convirtiéndose en un conjunto de patrones gestionados algorítmicamente. Se trata de un poder invisible y anticipatorio, adherido al diseño de la experiencia digital. Diversos autores han señalado que la GA no es mera teoría, sino una práctica extendida en instituciones, economías y plataformas. Rodríguez (2018) advierte que la recolección de metadatos ha transformado el gobierno, delegando decisiones en algoritmos que disminuyen la autonomía. Gorriti (2021) describe cómo plataformas como Google moldean conductas a través de personalización y publicidad dirigida, desplazando el poder hacia el diseño y la adherencia afectiva. Green (2022) alerta de que confiar solo en supervisión humana es ilusorio sin marcos democráticos fuertes. Cristianini y Scantamburlo (2020) analizan el crédito social en China como muestra de cómo la correlación estadística reemplaza la ley mediante premios y castigos.

Otros estudios señalan los desafíos institucionales. Jonk e Iren (2021), en municipios de Países Bajos, evidencian la falta de marcos unificados de gobernanza algorítmica, lo que genera decisiones opacas. Levy, Chasalow y Riley (2021) muestran que en justicia penal y educación en Estados Unidos se usan algoritmos sin estándares claros, perpetuando sesgos y minando legitimidad. En la vida cotidiana, Ouellet, Ménard, Bonenfant y Mondoux (2015) analizan el self-tracking como forma de autogobierno algorítmico. Perea Acevedo (2016)

conecta estas dinámicas con las “técnicas de sí” foucaultianas, subrayando la exigencia de optimización constante. Sadowski (2019) entiende los datos como un nuevo tipo de capital, lo que permite ver la GA como explotación económica de la vida digital. García Canclini (2020) denomina a este fenómeno “desciudadanización”, señalando que los algoritmos limitan la deliberación ciudadana y moldean formas de participación subordinadas.

La crítica descolonial advierte que la GA en el Sur global reproduce lógicas coloniales. Tironi y Valderrama (2021) muestran cómo los sistemas algorítmicos sostienen estructuras de dominación, pese a su aparente neutralidad. Cófreces (2021) enfatiza que la privatización de infraestructuras digitales y la opacidad algorítmica instauran un nuevo control social, donde la experiencia de los usuarios se dirige y monetiza constantemente. En síntesis, la racionalidad algorítmica no se limita a mejorar procesos: redefine el poder. La vigilancia distribuida, la regulación insuficiente, la explotación de datos y la desciudadanización digital convergen en un régimen que organiza la vida mediante anticipación y correlación. Este gobierno digital transforma los vínculos entre instituciones, plataformas y sujetos, tensiona la autonomía y la equidad, y exige nuevas formas de responsabilidad y resistencia. Las cuatro líneas estructurantes, ofrecen una cartografía crítica de estos vectores, mostrando cómo la GA atraviesa cuerpos, saberes y

prácticas en la configuración de un orden político-tecnológico emergente.

## **Líneas de fuga y expresiones actuales para entender la GA**

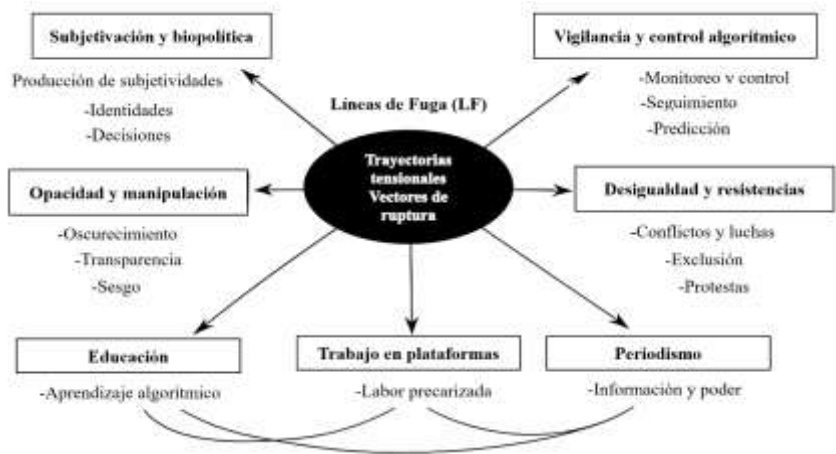
Tras la investigación realizada, y a partir de una revisión crítica de la literatura, se identificó que la GA opera a través de múltiples frentes que pueden sistematizarse mediante el concepto de líneas de fuga (LF). En el contexto de esta investigación, las LF no se entienden como categorías temáticas, áreas de impacto o simples vías de escape, pues, como afirman Deleuze y Guattari (2010), “las líneas de fuga no consisten nunca en huir del mundo, sino más bien en hacer que ese mundo huya, como cuando se agujerea un tubo” (p. 208). Por el contrario, se definen como trayectorias tensionales y vectores de ruptura en los que el poder algorítmico se despliega, se intensifica o se fractura.

Esta cartografía crítica permite evitar un mero inventario descriptivo, agrupando las diversas expresiones del poder y las resistencias que suscita en cuatro grandes líneas estructurantes: 1) Subjetivación y biopolítica; 2) Vigilancia y control algorítmico; 3) Opacidad, transparencia y manipulación; y 4) Desigualdad, exclusión y resistencias. En torno a estos cuatro vectores principales se subordinan y articulan dominios empíricos específicos, como la educación, el trabajo en plataformas y el periodismo, entendidos aquí

como los escenarios vitales donde dichas tensiones operan y se materializan.

**Figura 1**

*Cartografía analítica: vectores estructurantes y dominios empíricos de la GA.*



*Nota.* Elaboración propia.

*Primera línea estructurante: subjetivación, biopolítica y delegación de decisiones*

Este primer vector estructurante evidencia cómo la GA no solo administra datos poblacionales, sino que modela la vida misma, las identidades, los afectos y los modos de autocomprensión humana. Esta línea tensional absorbe

dinámicas biopolíticas, la explotación psicopolítica y la automatización de la agencia. Desde una matriz de origen foucaultiano, la gobernanza actual incorpora una dimensión biopolítica al gestionar poblaciones y cuerpos a través de tecnologías algorítmicas. Dencik (2022) sostiene que el capitalismo contemporáneo se estructura crecientemente en torno a la economía de los datos, transformando no solo el mercado, sino también la vida social y política, mientras Toscano López (2023) describe la convergencia entre biotecnología, inteligencia artificial (en adelante IA) y plataformas como la instauración de nuevos modos de biocontrol genético.

Sin embargo, este biopoder algorítmico avanza hacia un nivel mucho más íntimo: la formación de la subjetividad. Rodríguez (2018) advierte que la subjetivación contemporánea se produce a través de dispositivos que anticipan comportamientos y limitan la agencia de la persona. Esta ideología técnica, como concluyen García Ramírez y Valle Jiménez (2020), lleva al individuo a incorporar pasivamente estructuras de control y adoptar identidades prefiguradas por la lógica algorítmica. En este estado de hiperconexión, Barboza (2022) sostiene que la GA captura la atención y debilita cualquier capacidad de resistencia.

Es en este punto donde la biopolítica deviene psicopolítica, convirtiendo los afectos y las emociones en materia prima de explotación. Barrios-Tao y Díaz Pérez

(2024), junto con Niño Morales, Cabeza Herrera y Flórez Pabón (2024) —apoyados en Byung-Chul Han—, muestran cómo la vigilancia afectiva y la IA operan como mecanismos de manipulación emocional que modulan la conducta. En una clave aún más profunda, desde el psicoanálisis, autores como Ramé (2024) y Prueger (2024) coinciden en que el “lenguaje cibernético” coloniza incluso el inconsciente, reemplazando la lógica significante, neutralizando complejos psíquicos que podrían ser emancipadores y reorganizando los modos de goce y deseo de los sujetos.

La consecuencia final de esta trayectoria de subjetivación es la renuncia a la propia agencia mediante la delegación de decisiones. La reconfiguración del sujeto frente a la máquina deriva en una fe ciega hacia el sistema predictivo. Ayala-Coqui (2023) rastrea esta creencia hasta una gubernamentalidad propia de Silicon Valley que asume la racionalidad algorítmica como infalible. No obstante, como afirman Prueger (2024), Bazzara (2021) y Tejero (2020), delegar incesantemente decisiones a máquinas opacas reduce la capacidad crítica, erosiona la deliberación democrática y multiplica las injusticias cotidianas.

Para ilustrar cómo estas dinámicas de subjetivación y biopolítica intervienen en la cotidianidad, este trabajo recurre a casos empíricos que funcionan como viñetas ejemplificadoras. Una primera viñeta que evidencia esta captura psicopolítica es el experimento de contagio emocional



llevado a cabo por Facebook en 2014. En dicho estudio, la plataforma alteró deliberadamente el algoritmo del *feed* de noticias de 689.003 usuarios, demostrando que esta manipulación incidía de manera directa en sus estados de ánimo (Kramer, Guillory, & Hancock, 2014). Esta intervención confirma empíricamente cómo las emociones se han convertido en un nuevo espacio de control invisible.

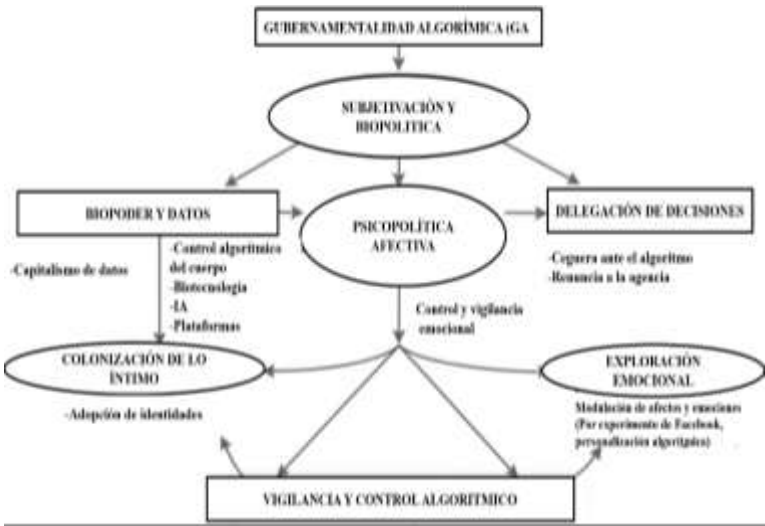
Una segunda viñeta se observa en la personalización algorítmica de plataformas como TikTok. Más allá de predecir gustos, su arquitectura de recolección de datos y refuerzo de burbujas de interés altera profundamente la autopercepción del usuario, moldeando la autopercepción y el valor que las personas se atribuyen, transformando profundamente lo que queremos hacer, quiénes deseamos ser y lo que pensamos de nosotros mismos (Bucher, 2018 citando a Hacking, 1990). En ambos casos, se hace visible que la subjetividad contemporánea, lejos de operar con autonomía, es sutilmente gobernada por una predeterminación algorítmica.

En suma, la captura psicopolítica del algoritmo demuestra que el sujeto digital no es verdaderamente libre cuando expresa sus emociones. Como advierten Deleuze y Guattari (2010), el sistema asegura su control porque “la subjetivación impone a la línea de fuga una segmentariedad que no cesa de renegar de ella, y a la desterritorialización absoluta un punto de abolición que no cesa de obstruirla, de desviarla” (p. 137). Sin embargo, esta captura íntima de la

subjetividad y la colonización de las emociones no serían posibles sin una infraestructura tecnológica capaz de monitorear y extraer información de cada movimiento en tiempo real. Es precisamente esta necesidad de rastreo continuo la que hace que la biopolítica se entrelace con un seguimiento inmanente, dando paso a la segunda línea estructurante de esta cartografía: la maquinaria de vigilancia masiva y el control algorítmico.

**Figura 2**

*Dinámicas de la primera línea estructurante: biopoder, captura psicopolítica y automatización de la agencia.*



Nota. Elaboración propia.

## *Segunda línea estructurante: vigilancia y control algorítmico*

El segundo vector estructurante expone cómo la GA opera mediante el seguimiento inmanente, continuo y anticipatorio. Este eje absorbe dinámicas de vigilancia masiva, la algoritmización del trabajo en plataformas y el colonialismo digital, conformando un sistema de control que va de la escala global a la microfísica de la cotidianidad. En su dimensión más abarcadora, la vigilancia algorítmica se describe como una modalidad anticipativa propia de la sociedad de metadatos (Rodríguez, 2018). Este vector se materializa en lo que Zuboff (2019) denomina “capitalismo de vigilancia”, cuyo emblema principal es el uso de datos por parte de corporaciones como Google para monetizar la experiencia humana (Gorriti, 2021). Sadowski (2019) profundiza en esta lógica al concebir los datos como un nuevo capital que inaugura dinámicas inéditas de explotación. Esta vigilancia masiva y distribuida amenaza como lo señala Tejero (2020), con derivar en un “ciber totalitarismo”. En consecuencia, la manipulación informativa amplifica el sensacionalismo y distorsiona la vida social (Almanza, 2021). Al mismo tiempo. Se instauran nuevas formas de biocontrol genético y de plataformas (Toscano López, 2023). En este marco, como apuntan Barrios-Tao y Díaz Pérez (2024), la extracción de datos emocionales perfecciona el control social, induciendo a lo que García Canclini (2020) llama la “desciudadanización” del individuo.

Para dar cuenta de cómo este vector tensional se inscribe en lo real, la cartografía despliega ahora, las escenas empíricas que actúan como viñetas, que no busca acumular pruebas definitivas, sino evocadoras de las sendas teóricas que se trazan. La primera viñeta de esta vigilancia anticipatoria es la predicción policial en Estados Unidos (Chicago y Los Ángeles). Como evidencian Brayne (2020), Eubanks (2021) y O'Neil (2016), los sistemas que evalúan riesgos de delitos o reincidencia se basan en datos históricos, convirtiendo la vigilancia en una máquina que perpetúa problemas raciales y socioeconómicos bajo una apariencia de eficiencia técnica. Otras viñetas operan a nivel del cuerpo y la intimidad, como lo plantea Ada Lovelace Institute (2020), desde el uso de aplicaciones de rastreo digital durante la pandemia de COVID-19, que disolvieron la privacidad en nombre de la seguridad colectiva, hasta la personalización predictiva de TikTok, que encasilla al usuario en burbujas de interés, afectando su autopercepción; y las prácticas de *self-tracking* o cuantificación personal, donde el individuo se convierte voluntariamente en el vigilante de sí mismo (Ouellet et al., 2015).

A partir de esta matriz de vigilancia, emergen dos derivaciones o subcampos críticos. El primero es la economía de plataformas y el trabajo algorítmico. En el ámbito laboral, siguiendo la lógica de Foucault (2009), el algoritmo actúa como un “nuevo soberano” que, no prohíbe directamente, sino que incentiva y sanciona de forma automática, y, desde

la lógica de Han (2021), induce a la autoexplotación del trabajador. Prassl (2018) y Castro Arghelets (2023) advierten cómo esta automatización redefine las relaciones laborales legales, mientras Noble (2018) evidencia que la aparente eficiencia de los sistemas algorítmicos está profundamente atravesada por dinámicas de vigilancia y discriminación. Investigaciones etnográficas, como las de Bergua, Montañés y Báez (2023) sobre los *riders*, junto con los análisis de Nava Becerra y Alvarado García (2023), confirman que la digitalización impone una profunda pérdida de autonomía. Dos viñetas empíricas ilustran claramente este subcampo: el modelo de Uber, donde los conductores están sometidos a sistemas automatizados que regulan precios y evalúan su rendimiento en tiempo real (Rosenblat, 2018). Y la automatización de recursos humanos en los almacenes de Amazon en Polonia, donde las métricas de productividad determinan la continuidad laboral sin intervención humana significativa (Rozmysłowicz & Krzyżaniak, 2023).

El segundo subcampo derivado de esta vigilancia estructurante es el colonialismo digital y extractivismo de datos. Esta lógica de control requiere de una infraestructura global profundamente asimétrica. Tello (2023) desarrolla el concepto de “colonialidad tecnológica del poder” para describir cómo el Sur Global depende de las corporaciones transnacionales. Este proceso es ratificado por García Canclini (2020), quien destaca cómo la exclusión digital reproduce lógicas coloniales, y por Couldry y Mejias (2020),

quienes advierten sobre la apropiación capitalista de la vida misma y sus altos costos de conexión. La viñeta ejemplificadora de esta tensión se observa en los países africanos, donde grandes corporaciones extraen masivamente datos para alimentar y entrenar sistemas de IA desarrollados en el Norte global, sin dejar ninguna inversión en infraestructura local, lo que consolida una nueva jerarquía de dominación informacional (Couldry & Mejias, 2020).

La economía de plataformas ilustra a la perfección el paso de la disciplina arquitectónica al control inmanente. En este escenario, el algoritmo encarna la tesis de Deleuze y Guattari (2010), en la cual:

cada centro de poder también es molecular, se ejerce sobre un tejido micrológico en el que ya sólo existe como difuso, disperso, desmultiplicado, miniaturizado, constantemente desplazado, actuando por segmentaciones finas, operado en el detalle, y en el detalle de detalles. (p. 228)

Ahora bien, la eficacia de esta vigilancia extractivista y del control incesante sobre los cuerpos —tanto en el trabajo de plataformas como a escala global— radica precisamente en su invisibilidad. Para que la extracción de datos opere sin resistencia inmediata, sus mecanismos de evaluación predictiva deben permanecer deliberadamente ocultos al escrutinio público, lo que nos conduce inevitablemente a la

tercera línea estructurante de esta cartografía: la “caja negra” de la opacidad algorítmica y los debates en torno a la transparencia.

**Figura 3**

*Derivaciones de la segunda línea estructurante: vigilancia inmanente, economía de plataformas y colonialismo digital.*



*Nota.* Elaboración propia.

### *Tercera línea estructurante: opacidad, transparencia y manipulación*

El tercer vector estructurante aborda la “caja negra” de la GA: la asimetría de información, los límites de la regulación ética y la capacidad del diseño técnico para manipular el discurso público. En esta línea convergen las tensiones normativas sobre la rendición de cuentas institucional y las derivaciones prácticas en el campo de la comunicación. El problema central de este eje radica en que la opacidad algorítmica refuerza el poder corporativo al tiempo que erosiona la equidad social y la participación cívica. Frente a esto, surge la demanda por una ética algorítmica y marcos regulatorios sólidos. Sin embargo, Green (2022) advierte que la presencia formal de supervisores humanos suele ser meramente simbólica y no garantiza la corrección de sesgos ni la transparencia, por lo que exige un verdadero control democrático y sustantivo.

Esta carencia regulatoria es evidenciada por Jonk e Iren (2021), quienes demuestran que ni siquiera en la administración pública se comprende el proceso algorítmico en su totalidad, dificultando enormemente la rendición de cuentas. En el ámbito de la justicia y los derechos fundamentales, autores como Levy et al., (2021), Ortega Klein (2025) y Huergo Lora (2025) alertan que confiar en la automatización opaca —como la figura del “juez robot”— perpetúa desigualdades, por lo que resulta imperativo exigir



explicabilidad institucional, responsabilidad compartida e involucrar a la sociedad civil en el debate sobre los fines de la IA, tal como también proponen Cristianini y Scantamburlo (2020) y Burgos y Hernández (2021).

La consecuencia más visible de esta opacidad se materializa en la manipulación algorítmica de la esfera pública. Aguilera, Sosa y De Aguilera (2018) sostienen que la automatización de la comunicación, al carecer de un acceso democrático, refuerza estructuralmente los discursos hegemónicos. En este sentido, las plataformas privilegian contenidos de alto impacto emocional que promueven la polarización y la distorsión de la vida social (Almanza, 2021). Del mismo modo, perfilan lo que Tejero (2020) denomina un totalitarismo digital basado en la propaganda encubierta. Como subcampo y área de impacto directo de esta tensión aparece el periodismo de datos. Flores Vivar (2018) explica que, si bien la digitalización exige nuevas competencias analíticas, depender excesivamente de algoritmos opacos subordina la profundidad informativa a la inmediatez mecánica.

Con el propósito de evidenciar la incidencia de este vector tensional en la realidad, la cartografía despliega las viñetas que ilustran los límites y desafíos de la llamada caja negra algorítmica. La primera se sitúa en el terreno de la regulación: en la Unión Europea, la Ley de Inteligencia Artificial exige supervisión humana y explicabilidad en

sistemas de alto riesgo (European Commission, 2021). A ello se suma la creación de registros públicos de algoritmos en ciudades como Ámsterdam y Helsinki, con el fin de que los ciudadanos comprendan los sistemas en uso y sus propósitos, facilitando la exigencia de responsabilidad (AlgorithmWatch, 2020). Estas iniciativas buscan mitigar la opacidad, aunque persiste el desafío de evitar que la supervisión se reduzca a un mero procedimiento burocrático.

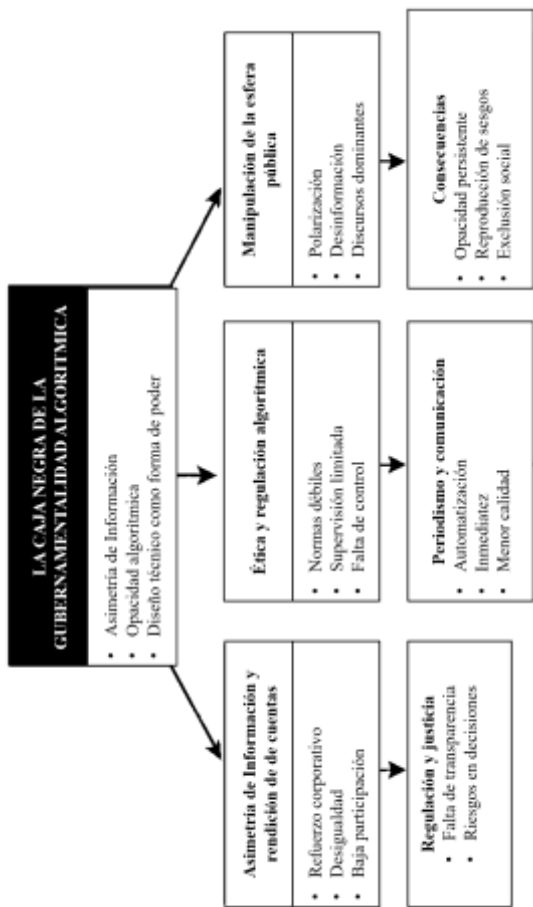
Una segunda viñeta ejemplifica la manipulación a gran escala: las elecciones presidenciales de Estados Unidos en 2016. En dicho contexto, el algoritmo de plataformas como Facebook amplificó burbujas informativas, permitiendo que noticias falsas a favor de un candidato fueran compartidas 30 millones de veces, demostrando que la arquitectura técnica no es ecuánime, sino que estructura activamente la polarización política (Allcott & Gentzkow, 2017). Finalmente, una tercera viñeta se encuentra en la automatización de noticias en agencias de Alemania, donde el software redacta reportes deportivos y financieros mecánicamente; este caso evidencia que, aunque la tecnología optimiza el volumen de cobertura, requiere forzosamente de edición humana para dotar a la información de contexto, variedad y calidad (Kunert, 2020). De esta manera, revela los límites éticos y prácticos de ceder la comunicación al cálculo algorítmico.

Este régimen de opacidad confirma que la GA opera fuera de los marcos representacionales humanos,

consolidándose como un ensamblaje ciego. En términos de Deleuze y Guattari (2010), la “máquina abstracta no es más física o corporal que semiótica, es *diagramática* (ignora tanto más la distinción entre lo natural y lo artificial). Actúa por *materia*, y no por sustancia; por *función*, y no por forma” (p. 144). Precisamente, es esta asimetría absoluta y la naturaleza opaca de sus funciones las que perpetúan los sesgos del sistema, expulsando a los sujetos más vulnerables hacia los bordes y dando paso a la cuarta línea estructurante de esta cartografía: la maquinaria de exclusión y las tensiones con los polos de resistencia.

**Figura 4**

*Tercera línea estructurante: la "caja negra" tecnológica, asimetría normativa y manipulación informativa.*



*Nota.* Elaboración propia.

#### *Cuarta línea estructurante: desigualdad, exclusión y resistencias*

El cuarto y último vector estructurante evidencia cómo el cálculo de lo probable perpetúa brechas históricas, al tiempo que suscita nuevos frentes de resistencia política y pedagógica. Esta línea de tensión absorbe el impacto social de los sesgos algorítmicos y ubica a la educación como un subcampo de exclusión, pero contrapone a estas lógicas la descolonización de los datos y el diseño crítico como herramientas de emancipación.

En su dimensión de exclusión, la GA reproduce y profundiza las desigualdades estructurales. Levy et al., (2021) señalan que en ámbitos determinantes los sesgos estadísticos perpetúan marginaciones preexistentes. Esta concentración del poder algorítmico en pocas corporaciones, como lo plantea Valente (2020) alimenta una precarización que afecta mayoritariamente a sectores laborales sin regulación. En clave latinoamericana, García Canclini (2020) advierte que las plataformas digitales exacerban la exclusión social, mientras Tello (2023) denuncia que la asimetría tecnológica condena al Sur Global a una profunda dependencia.

Estas brechas se agudizan en contextos informativos, donde la manipulación afecta de forma desproporcionada a colectivos vulnerables (Burgos & Hernández, 2021). Además, fomenta la polarización social (Almanza, 2021). Un subcampo

donde esta desigualdad se materializa con fuerza es la educación. Bañares-Marivela (2024) analiza cómo la intermediación algorítmica y la lógica comercial de las plataformas desplazan los discursos pedagógicos inclusivos, imponiendo métricas de rendimiento que marginan la diversidad cognitiva. Frente a esta maquinaria de exclusión, la GA genera un polo de resistencia articulado en torno a las epistemologías del Sur y el diseño crítico. Como alternativa a la reproducción de dinámicas coloniales, según García Canclini (2020) y Tello (2023), emerge la necesidad de alfabetizaciones críticas y soberanías digitales. Tironi y Valderrama (2021) sostienen que el diseño crítico y la pedagogía descolonial son claves para visibilizar y subvertir la clasificación algorítmica, especialmente en contextos vulnerables.

Para ilustrar las tensiones y manifestaciones de estas dinámicas de opresión y resistencia, las viñetas que dejan ver su operatividad en esta cartografía se exhiben en seguida: En cuanto a la amplificación de la desigualdad, una primera viñeta se observa en la automatización de programas de asistencia social en Indiana (EE. UU.). Allí, un sistema automatizado canceló masivamente beneficios por supuestas inconsistencias sin considerar factores contextuales, lo que demuestra que la automatización sin supervisión convierte al Estado en un aparato de injusticia ciega (Eubanks, 2021). Una segunda viñeta ilustra la exclusión en el subcampo educativo: el proctoring automatizado (sistemas de vigilancia de

exámenes). Se ha comprobado que estos algoritmos presentan sesgos raciales y de género intolerables, marcando como sospechosos y fallando en el reconocimiento facial de estudiantes con tonos de piel oscuros, particularmente mujeres negras (Koencke et al., 2020).

Por otro lado, como viñetas ejemplificadoras de las resistencias y la descolonización, destaca el caso de la soberanía de datos indígenas en comunidades de Australia y Nueva Zelanda. Estas comunidades reivindican el control sobre la recolección y gestión de su información, oponiéndose a las lógicas extractivistas mediante un acto político que defiende la dignidad cultural (Kukutai & Taylor, 2016). De igual modo, como lo expone Kibaroglu, (2020), las cripto finanzas y el blockchain han abierto debates sobre la gestión de identidades digitales soberanas sin mediadores, aunque conllevan el riesgo de crear nuevas exclusiones si no se reestructura el poder (Couldry & Mejias, 2020).

Finalmente, una última viñeta de pedagogía descolonial en acto se encuentra en los hackerspaces brasileños. Organizados en colectivos barriales, desarrollan tecnologías comunitarias y alfabetización digital crítica, demostrando —al igual que prototipos artísticos como *Big Sister*— que el diseño de software puede ser subvertido para construir futuros tecnológicos arraigados en la justicia colectiva (Martins & Albagli, 2020; Tironi & Valderrama, 2021). A pesar de la contundencia de esta maquinaria de

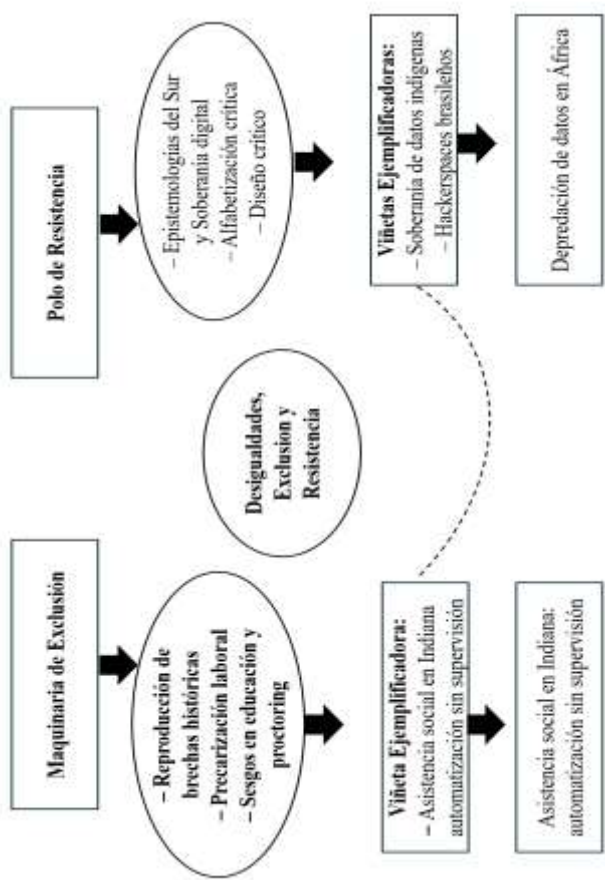
exclusión, las prácticas de soberanía tecnológica demuestran que el control total es ilusorio.

Frente al cerco algorítmico, la agencia humana y la resistencia comunitaria emergen con fuerza porque, tal como afirman Deleuze y Guattari (2010), “un rizoma puede ser roto, interrumpido en cualquier parte, pero siempre recomienza según esta o aquella de sus líneas, y según otras. Es imposible acabar con las hormigas” (p. 15). De este modo, la fuga vuelve inevitablemente a abrirse paso. Sin embargo, para comprender la verdadera magnitud de la GA, es preciso reconocer que estas cuatro trayectorias no operan de forma aislada, sino que convergen y se retroalimentan, lo que nos conduce al siguiente apartado: las articulaciones estructurales y la consolidación de la economía de plataformas como epicentro tensional.



**Figura 5**

*Cuarta línea estructurante: tensiones entre la maquinaria de exclusión y los polos de resistencia emancipadora.*



*Nota.* Elaboración propia.

## **Articulaciones estructurales: la economía de plataformas como epicentro tensional**

El recorrido realizado hasta aquí permite evidenciar que las cuatro líneas estructurantes de la GA no operan de forma aislada, sino que se retroalimentan, conformando una maquinaria de poder integral. En este sentido, la economía de plataformas funciona bajo la lógica de lo que Deleuze y Guattari (2010) definen como un ensamblaje complejo: un “*agenciamiento maquínico* de cuerpos, de acciones y de pasiones, mezcla de cuerpos que actúan los unos sobre los otros” (p. 92). Al utilizar algoritmos predictivos para asignar, evaluar y monetizar el trabajo o la educación, estas infraestructuras trascienden lo meramente económico para convertirse en verdaderos laboratorios de la racionalidad algorítmica.

La articulación comienza con el despliegue de la Vigilancia y el Control Algorítmico (Segunda LF). Plataformas como Uber o los sistemas de aprendizaje en línea estandarizan la actividad a través de cálculos automatizados que fijan objetivos, miden la eficiencia y clasifican a los individuos en tiempo real. Esta vigilancia inmanente opera bajo un régimen estricto de opacidad, transparencia y manipulación (tercera LF). Como señalan García Canclini (2020) y Tello (2023), la falta de transparencia y la discrecionalidad en el diseño de estos sistemas impiden que el trabajador o el estudiante comprendan las reglas bajo las

cuales son medidos, lo que consolida al algoritmo como un “nuevo soberano” que, siguiendo a Foucault (2009), no necesita prohibir de manera directa, sino que orienta el comportamiento mediante recompensas y sanciones opacas.

Es precisamente la convergencia entre vigilancia incesante y opacidad estructural lo que detona la subjetivación, biopolítica y delegación de decisiones (Primera LF). Al ignorar los criterios exactos de evaluación, pero saberse continuamente observado, el sujeto interioriza el control. Como advierte Byung-Chul Han (2021), los sistemas digitales captan la atención y el comportamiento de las personas llevándolas a dinámicas de autoexplotación. En las plataformas de reparto, el trabajador altera sus propias rutinas, emociones y tiempos de descanso para asegurar “buenas calificaciones”, demostrando un poder que coloniza la interioridad y modela la vida misma (Toscano López, 2023).

Finalmente, este ciclo de vigilancia, opacidad y autoexplotación cristaliza en la Desigualdad, Exclusión y Resistencias (Cuarta LF). El sistema algorítmico produce nuevas capas de segregación social, operando como aparatos de estratificación que, en términos de Deleuze y Guattari (2010), funcionan como verdaderas capturas, “como 'agujeros negros' u oclusiones que se esforzaban en retener todo lo que pasaba a su alcance” (p. 48). La mezcla de estos factores técnicos y económicos acentúa las brechas históricas,

excluyendo a quienes carecen de infraestructura tecnológica o penalizando la diversidad de pensamiento en entornos educativos dominados por la lógica de la rentabilidad (Baker & Hawn, 2022). Sin embargo, es también en este punto de máxima tensión donde emergen las líneas de resistencia: Frente a la precarización y la manipulación algorítmica, se hace urgente articular soberanías digitales y epistemologías críticas que fracturen el control de las plataformas.

En resumen, la GA opera mediante la articulación simultánea de estos cuatro vectores: vigila para extraer valor, oculta sus mecanismos para evitar rendir cuentas, subjetiva para inducir obediencia automática, y estratifica a la sociedad perpetuando la desigualdad.

**Figura 6**

*Articulación estructural: la economía de plataformas como núcleo convergente de la racionalidad algorítmica.*



*Nota.* Elaboración propia.

**Consideraciones finales: gubernamentalidad algorítmica en acto y horizontes por pensar**

Más que un cierre definitivo, estas consideraciones finales proponen una recapitulación crítica de lo expuesto, señalando a la vez caminos posibles de investigación futura. La GA, tal como se ha mostrado, no es únicamente un despliegue técnico, sino un proyecto político de largo alcance

que transforma las instituciones redefine las subjetividades e incide en las formas de vida. La digitalización de lo cotidiano, junto con la anticipación estadística de las conductas, convierte a los algoritmos en operadores de gobierno que modulan desde la intimidad hasta la esfera pública.

Las cuatro líneas estructurantes sistematizadas en esta cartografía permiten comprender que la GA actúa como un entramado de control, delegación y explotación, pero también como un campo donde se manifiestan tensiones y resistencias:

En primer lugar, la actualización digital de la subjetivación y la biopolítica muestra cómo la vida se gobierna a través de rastreos epidemiológicos, registros emocionales y la delegación de decisiones. La soberanía corporal y la autonomía psíquica se ven desafiadas por una lógica de captura que penetra la interioridad. Esto exige repensar filosóficamente la relación entre cuerpo, datos y poder: ¿qué significa la soberanía en un mundo donde el organismo se convierte en interfaz algorítmica? Y, de manera más profunda, ¿hasta dónde puede externalizarse la agencia humana sin que se erosione la condición ontológica de deliberar?

En segundo lugar, la vigilancia y el control algorítmico, expresados crudamente en la economía de plataformas y el colonialismo digital, evidencian un régimen de precarización

en el que la autonomía se diluye en la necesidad de rendimiento continuo. El Sur Global es a menudo tratado como proveedor de información y usuario subordinado de infraestructuras de extracción. El reto aquí no es solo denunciar las condiciones de explotación, sino imaginar plataformas cooperativas y ecosistemas tecnológicos diseñados bajo principios de justicia redistributiva.

En tercer lugar, el vector de la opacidad, transparencia y manipulación revela que los marcos normativos aún se encuentran rezagados frente a las innovaciones tecnológicas; incluso allí donde existen propuestas avanzadas, como la Ley de IA en Europa o los registros públicos, la dificultad radica en garantizar que la supervisión no sea meramente formal. El gobierno de la atención y la explotación de la afectividad constituyen vectores de control invisibles. Queda abierta la pregunta: ¿cómo asegurar que los marcos jurídicos acompañen críticamente la velocidad del cambio tecnológico y no se limiten a responder *ex post facto*?

Finalmente, frente a la desigualdad y exclusión que estos sistemas perpetúan al basarse en datos impregnados de sesgos históricos, emergen con fuerza las resistencias críticas. Las perspectivas de descolonización, el diseño crítico y las iniciativas de soberanía de datos indígenas sugieren que la lucha tecnológica no se libra solo en términos de eficiencia, sino también de territorialidad, cultura y cosmovisión. De aquí surge una cuestión crucial para el futuro: ¿qué prácticas

comunitarias y pedagógicas pueden generar racionalidades tecnológicas alternativas, que no solo mitiguen sesgos, sino que reimaginen formas de inclusión desde abajo?

En síntesis, las líneas de fuga aquí analizadas expresan procesos simultáneos de desterritorialización y reterritorialización. La GA configura nuevas prisiones de datos, pero, en el mismo movimiento, detona fisuras conceptuales que habilitan la creación de nuevas posibilidades y agenciamientos colectivos para la emancipación humana.

## Referencias

- Ada Lovelace Institute. (2020). *Exit through the App Store? Rapid evidence review on the technical considerations and societal implications of using technology to transition from the COVID-19 crisis*.  
<https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-Rapid-Evidence-Review-Exit-through-the-App-Store-April-2020-2.pdf>
- Aguilera, M., Sosa, A., & de Aguilera, R. (2018). Comunicación, discursos, algoritmos, poder. *Ámbitos. Revista Internacional de Comunicación*, 40.  
<https://idus.us.es/items/ada15715-edfa-461e-9d7d-fdbe7eed6ef6>



- AlgorithmWatch. (2020). *Automating society report 2020*. Bertelsmann Stiftung.  
<https://algorithmwatch.org/en/project/automating-society-2020/>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236.  
<https://doi.org/10.1257/jep.31.2.211>
- Almanza, A. (2021). El poder del algoritmo y la vida social. *Revista de Estudios Sociales*, 71, 274.
- Amoore, L. (2020). *Cloud ethics*. Duke University Press.
- Ayala-Colqui, J. (2023). El nacimiento del "ciberalismo": Una genealogía crítica de la gubernamentalidad de Silicon Valley. *Bajo Palabra*, 32, 221–254.  
<https://doi.org/10.15366/bp2023.32.012>
- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092.  
<https://doi.org/10.1007/s40593-021-00285-9>
- Bañares-Marivela, E. (2024). El poder de los algoritmos en la mercantilización de las redes sociales para la educación inclusiva: ¿Invisibles o invisibilizados? En R. Roig-Vila, J. M. Antolí Martínez, C. Lorenzo Álvarez, V. Onrubia-Martínez, & A. Sánchez Ronco (Eds.), *Libro de actas. 3rd International Congress: Humanities and Knowledge* (p. 15). Ediciones Octaedro.

- Barboza, R. A. (2022). Gubernamentalidad algorítmica y subjetividad. *Textos y Contextos*, 1(24), Artículo e3483. <https://doi.org/10.29166/tyc.v1i24.3483>
- Barrios-Tao, H., & Díaz, V. (2024). Inteligencia artificial y emociones: Psicopolítica mediante datos y algoritmos (2015-2022). *Revista de Ciencias Sociales*, 30(1), 251–267.
- Bazzara, L. (2021). ¿Todo el poder a los algoritmos? Asistencias, delegaciones y modulaciones en la nueva razón gubernamental. *Argumentos. Revista de Crítica Social*, 23, 33–58.
- Bergua, J. Á., Montañés, M., & Báez, J. M. (2023). Riders, algoritmos y negociación de la realidad. *Empiria. Revista de Metodología de Ciencias Sociales*, 59, 81–102. <https://doi.org/10.5944/empiria.59.2023.37939>
- Brayne, S. (2020). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press. <https://doi.org/10.1093/oso/9780190684099.001.0001>
- Bucher, T. (2018). *If...Then: Algorithmic power and politics*. Oxford University Press. <https://doi.org/10.1093/oso/9780190493028.001.0001>
- Burgos, E., & Hernández Díaz, G. (2021). El poder social de los algoritmos: ¿Ciudadanos desprotegidos? En M. Bisbal & G. Hernández (Eds.), *Descentramiento de la comunicación* (pp. 75–90). Centro Gumilla.

- Castro Arghelets, M. A. (2023). El ejercicio de los poderes empresariales mediante algoritmos. *Revista Jurídica de Asturias*, 45, 125–155.
- Cófreces, J. (2021, junio). *Los algoritmos no mienten: La articulación entre saber, poder y subjetividad en tiempos de la gubernamentalidad algorítmica* [Ponencia]. XIII Jornadas de Sociología, Facultad de Ciencias Sociales, Universidad de Buenos Aires, Argentina. <https://cdsa.aacademica.org/000-074/60>
- Couldry, N., & Mejias, U. A. (2020). *The costs of connection: How data are colonizing human life and appropriating it for capitalism*. Oxford University Press.
- Cristianini, N., & Scantamburlo, T. (2020). On social machines for algorithmic regulation. *AI & Society*, 35(3), 645–662. <https://doi.org/10.1007/s00146-019-00917-8>
- Deleuze, G., & Guattari, F. (2010). *Mil mesetas: Capitalismo y esquizofrenia* (J. Vázquez Pérez, Trad.; 6.<sup>a</sup> ed.). Editorial Pre-Textos. (Obra original publicada en 1980)
- Dencik, L. (2022). The datafied welfare state: A perspective from the UK. En A. Hepp, J. Jarke, & L. Kramp (Eds.), *New perspectives in critical data studies: Transforming communications – Studies in cross-media research* (pp. 165–183). Palgrave Macmillan. [https://doi.org/10.1007/978-3-030-96180-0\\_7](https://doi.org/10.1007/978-3-030-96180-0_7)

- Eubanks, V. (2021). *La automatización de la desigualdad: Cómo las herramientas de alta tecnología perfilan, vigilan y castigan a los pobres* (G. De Aza, Trad.). Capitán Swing Libros.
- European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)* (COM/2021/206 final). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- Flores-Vivar, J. M. (2018). Algoritmos, aplicaciones y Big data, nuevos paradigmas en el proceso de comunicación y de enseñanza-aprendizaje del periodismo de datos. *Revista de Comunicación*, 17(2), 268–291. <https://doi.org/10.26441/RC17.2-2018-A12>
- Foucault, M. (2009). *Nacimiento de la biopolítica: Curso en el Collège de France (1978-1979)* (H. Pons, Trad.). Akal. (Obra original publicada en 2004)
- García Canclini, N. (2020). *Ciudadanos reemplazados por algoritmos*. FLACSO Ecuador.
- García Ramírez, D., & Valle Jiménez, D. (2020). Los impactos de la ideología técnica y la cultura algorítmica en la sociedad: Una aproximación crítica. *Revista de Estudios Sociales*, 71, 15–27. <https://doi.org/10.7440/res71.2020.02>

- Gorriti, J. (2021). Google y la gubernamentalidad algorítmica: Datos, relaciones e individualización. *Avatares de la Comunicación y la Cultura*, 22.
- Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45, 105681. <https://doi.org/10.1016/j.clsr.2022.105681>
- Han, B.-C. (2021). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder* (A. Bergés, Trad.; 2.<sup>a</sup> ed.). Herder Editorial.
- Huergo Lora, A. (2025). El uso de algoritmos por las administraciones públicas. En M. E. Casas Barrondo & D. Pérez del Prado (Eds.), *Derecho y tecnología*. Editorial Troia.
- Jonk, E., & Iren, D. (2021). Governance and communication of algorithmic decision making: A case study on public sector. *IEEE 26th Conference on Business Informatics (CBI)*.
- Kibaroglu, O. (2020). Self sovereign digital identity on the blockchain: A discourse analysis. *Financial Markets, Institutions and Risks*, 4(2), 65–79. [https://doi.org/10.21272/fmir.4\(2\).65-79.2020](https://doi.org/10.21272/fmir.4(2).65-79.2020)
- Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Touns, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National*

- Academy of Sciences*, 117(14), 7684–7689.  
<https://doi.org/10.1073/pnas.1915768117>
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), 8788–8790.  
<https://doi.org/10.1073/pnas.1320040111>
- Kukutai, T., & Taylor, J. (2016). Data sovereignty for indigenous peoples: Current practice and future needs. En T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (Vol. 38, pp. 1–22). ANU Press. <http://www.jstor.org/stable/j.ctt1q1crgf.8>
- Kunert, J. (2020). Automation in sports reporting: Strategies of data providers, software providers, and media outlets. *Media and Communication*, 8(3), 5–15.  
<https://doi.org/10.17645/mac.v8i3.2996>
- Levy, K., Chasalow, K. E., & Riley, S. (2021). Algorithms and decision-making in the public sector. *Annual Review of Law and Social Science*, 17, 309–334.  
<https://doi.org/10.1146/annurev-lawsocsci-041221-023808>
- Martins, B. C., & Albagli, S. (2020). Hackerspace movement: A study of the Brazilian experience. *Em Questão*, 26(1), 161–185. <https://doi.org/10.19132/1808-5245261.161-185>

- Nava Becerra, M. E., & Alvarado García, V. M. (2023). Without the need for anyone (except for the algorithm). *Desde el Sur*, 15(2), e0020. <https://doi.org/10.21142/DES-1502-2023-0020>
- Niño Morales, E. E., Cabeza Herrera, O. J., & Flórez Pabón, C. E. (2024). Psicopolítica y big data como nuevas formas y herramientas para la organización política. *Sophia, Colección de Filosofía de la Educación*, 36, 247–273. <https://doi.org/10.17163/soph.n36.2024.08>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Ortega Klein, A. (2025). Juez con robot, más que robot juez: Asistencias, delegaciones y modulaciones en la nueva razón gubernamental. En M. E. Casas Baamonde (Dir.) & D. Pérez del Prado (Coord.), *Derecho y tecnologías* (pp. 45–75). Editorial Centro de Estudios Ramón Areces.
- Ouellet, M., Ménard, M., Bonenfant, M., & Mondoux, A. (2015). Big Data et quantification de soi: La gouvernementalité algorithmique dans le monde numériquement administré. *Canadian Journal of Communication*, 40(4), 597–613. <https://doi.org/10.22230/cjc.2015v40n4a2869>

- Perea Acevedo, A. J. (2016). *Michel Foucault: Vocabulario de nociones espaciales* (1.<sup>a</sup> ed.). Universidad Distrital Francisco José de Caldas; CLACSO; Editorial Magisterio.
- Prassl, J. (2018). *Humans as a service: The promise and perils of work in the gig economy*. Oxford University Press.
- Prueger, J. (2024). *Entre arquetipos y algoritmos: Las teorías del poder posfoucaultianas en hibridación con la psicología analítica de Carl G. Jung* (1.<sup>a</sup> ed.) [Libro digital, PDF]. Fundación CICCUS.
- Ramé, I. (2024). *El sujeto, preocupado por la biopolítica del algoritmo* [Ponencia]. XVI Congreso Internacional de Investigación y Práctica Profesional en Psicología. <https://www.aacademica.org/000-048/411.pdf>
- Rodríguez, P. (2018). Gubernamentalidad algorítmica: Sobre las formas de subjetividad en la sociedad de los mediados. *Revista Barba, Vida*, 4(Nex 6).
- Rodríguez, P. E. (2019). Gubernamentalidad algorítmica. En S. Puente (Ed.), *Las palabras y en las cosas: Saber, poder y subjetividad entre algoritmos y biomoléculas* (pp. 353–361). Editorial Catus.
- Rosenblat, A. (2018). *Uberland* (1.<sup>a</sup> ed.). University of California Press.
- Rouvroy, A. (2020, 27 de marzo). Algorithmic governmentality and the death of politics. *Green European Journal*.



<https://www.greeneuropeanjournal.eu/algorithmic-governmentality-and-the-death>

- Rouvroy, A., & Berns, T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation. *Réseaux*, 177(1), 163–196. <https://doi.org/10.3917/res.177.0163>
- Rozmysłowicz, M., & Krzyżaniak, P. (2023). Automated processing of data on work performance and employee evaluation: A case study of practices at Amazon warehouses in Poland. *Italian Labour Law e-Journal*, 16(2), 149–164. <https://doi.org/10.6092/issn.1561-8048/18084>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1), 1–12. <https://doi.org/10.1177/2053951718820549>
- Tejero, E. L. (2020). Algoritmos. El totalitarismo determinista que se avecina: ¿La pérdida final de libertad? *Revista de Pensamiento Estratégico y Seguridad CISDE*, 5(1), 85–101.
- Tello, A. (2023). Sobre el colonialismo digital: Datos, algoritmos y colonialidad tecnológica del poder en el sur global. *Inmediaciones de la Comunicación*, 18(2), 89–110. <https://doi.org/10.18861/ic.2023.18.2.3523>
- Tironi, M., & Valderrama, M. (2021). Descolonizando los sistemas algorítmicos: Diseño crítico para la problematización de algoritmos y datos digitales desde

el sur. *Palabra Clave*, 24(3), Artículo e2432.

<https://doi.org/10.5294/pacla.2021.24.3.2>

Toscano López, D. (2023). Biopolítica molecular y producción artificial de la vida: Dispositivos de biogeneración e infogeneración. *En-Claves del Pensamiento*, 34, e566.

<https://doi.org/10.46530/ecdp.voi34.566>

Valente, J. (2020). Apresentação do dossiê temático "Algoritmos, economia e poder". *Revista Eletrônica Internacional de Economia Política da Informação, da Comunicação e da Cultura*, 22(2), 57–63.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

# Mecanismos de participación ciudadana en la sociedad digital

*Mechanisms for Citizen Participation in the Digital Society*

**María Fernanda Idrobo Flores<sup>44</sup>, Boris Raúl Ochoa Ordoñez<sup>45</sup>**

## Resumen

La consolidación de la sociedad digital ha modificado profundamente las formas tradicionales de participación ciudadana, impulsando nuevos escenarios para la deliberación pública, la incidencia política y el control social. Este artículo analiza los principales mecanismos de participación ciudadana en el contexto digital, examinando su

---

<sup>44</sup> Universidad Técnica Particular de Loja. <https://orcid.org/0000-0002-2854-4056>

<sup>45</sup> Universidad Técnica Particular de Loja. <https://orcid.org/0000-0003-4928-0915>

evolución, alcances y limitaciones frente a los desafíos contemporáneos de la democracia. A partir de una revisión teórica y normativa, se aborda cómo las tecnologías de la información y la comunicación (TIC) han ampliado las posibilidades de interacción entre ciudadanía y Estado, promoviendo una participación más abierta, inmediata y descentralizada. Sin embargo, persisten obstáculos estructurales vinculados con la brecha digital, la falta de alfabetización tecnológica y la escasa institucionalización de los espacios digitales de participación. El estudio concluye que la participación ciudadana en la era digital constituye una oportunidad para fortalecer la democracia deliberativa, siempre que se garantice el acceso equitativo a las tecnologías, la transparencia en los procesos digitales y la protección de los derechos fundamentales en línea.

**Palabras clave:** Participación ciudadana; sociedad digital; gobernanza digital; democracia participativa.

## **Abstract**

The consolidation of the digital society has profoundly modified traditional forms of citizen participation, fostering new scenarios for public deliberation, political influence, and social oversight. This article analyzes the main mechanisms of citizen participation in the digital context, examining their evolution, scope, and limitations in the face of contemporary democratic challenges. Based on a theoretical and normative

review, it addresses how information and communication technologies (ICTs) have expanded the possibilities for interaction between citizens and the state, promoting more open, immediate, and decentralized participation. However, structural obstacles persist, linked to the digital divide, a lack of technological literacy, and the limited institutionalization of digital spaces for participation. The study concludes that citizen participation in the digital age represents an opportunity to strengthen deliberative democracy, provided that equitable access to technologies, transparency in digital processes, and the protection of fundamental rights online are guaranteed.

**Keywords:** Citizen participation; digital society; digital governance; participatory democracy.

## **Introducción**

En la última década, dentro del campo de estudio de la participación política se evidencia un incremento del interés por otros tipos de participación que inciden en la vida social, y que incluyen tanto factores políticos como no políticos (Delfino & Zubieta, 2010). A estas formas se las denomina participación ciudadana, misma que integra modalidades políticas convencionales y no convencionales, y otras como la vinculación con asociaciones comunitarias y ONGs. Para que estas acciones sean participativas deben ser realizadas por los individuos en su calidad de ciudadanos en ámbitos públicos,

y no como funcionarios que de una u otra manera tengan mayor afinidad al partido de turno. Así también, la participación es entendida como acción y ésta debe ser voluntaria (van Deth, 2001). Es importante detallar el enfoque dentro de la participación política, la distinción más habitual suele establecerse entre convencional y no convencional, la primera incluye todas aquellas acciones relacionadas con el proceso electoral, en tanto que la segunda integra el hacer peticiones, manifestaciones legales, huelgas ilegales, daño a la propiedad y violencia personal que evidentemente va en contra de los principios de libertad de expresión y asociación (Delfino, 2013). En cuanto a aquellas modalidades de participación con incidencia en la vida social pero no políticas, es decir, que no intentan tener injerencia en el curso de procesos político-sociales, y aquellas formas de participación ciudadana que se derivan de un contacto en particular. Banda Castro et al. (2010), a partir de una investigación, describen tres campos de acción de la participación ciudadana en las cuales se manifiestan los comportamientos de las personas: (a) participación ciudadana privada, vinculada al cumplimiento de deberes y obligaciones como el pago de impuestos; (b) participación ciudadana social, que se enfoca en mejorar las condiciones de vida de una determinada colectividad y se evidencia en juntas de la comunidad, sindicatos, grupos ecológicos, sociedades o colegios profesionales; y (c) participación ciudadana política, que es la que se realiza buscando canales de comunicación,

interacción e integración con instituciones estatales para lograr ciertas decisiones gubernamentales, todo esto relacionado con la comunicación digital.

### *Relevancia de la participación ciudadana en la era digital*

El desarrollo de las tecnologías digitales, la expansión de internet y el crecimiento de las redes móviles han generado profundos cambios en las formas políticas a nivel global y más aún cuando se lo vincula con la digitalización. De acuerdo con Vercelli (2007), las arquitecturas de participación ciudadana abiertas, libres y de producción colaborativa entre pares han permitido que, como nunca antes en la historia de la humanidad, millones y millones de personas (usuarios finales, ciudadanos) a nivel global puedan informarse, comunicarse, expresar sus preferencias, gestionar parte de sus derechos ciudadanos y colaborar para producir, disponer y compartir todo tipo de bienes y obras intelectuales. Estas tecnologías están cambiando la manera de hacer política: desde las campañas y sus diferentes formas de financiamiento, hasta las mismas bases de la participación política ciudadana.

En ningún caso estas tecnologías han sido neutras. Por un lado, han renovado, fortalecido y pueden enriquecer aún más el ejercicio directo e inmediato de derechos ciudadanos fundamentales, tales como: la libertad de expresión, el derecho a peticionar a las autoridades, el derecho de autor,

derecho de copia, el derecho de libre acceso y la disponibilidad de la información pública (Benkler, 2006; Lessig, 2001; Vercelli, 2009). Por otro lado, lejos de una neutralidad tecnológica, estas mismas tecnologías también han servido para generar nuevas formas de censura, menoscabos a derechos intelectuales, la desaparición de la privacidad, formas autoritarias de control, la privatización de informaciones de carácter público o la formación de gobiernos totalitarios (Lessig, 2008; Vercelli, 2009).

De esta manera, es que la relación entre las tecnologías digitales y las formas de gobierno ya tiene una larga historia, tomando en cuenta las promesas y reivindicaciones de mejores formas de gobierno, que estuvieron presentes en los inicios de la digitalización, desde la informatización de las administraciones públicas, hasta los más sofisticados y ambiciosos proyectos de gobiernos electrónicos (Berners-Lee, 2000; Castells, 2001; Naser & Concha, 2012). En la actualidad tal vez con más escepticismo, exigencias y urgencias, todavía sigue en pie la pretensión de construir gobiernos más democráticos, con una ciudadanía digital fuerte y, en suma, lograr una nueva y mejor relación entre los ciudadanos y sus gobernantes (Jorquera Beas, 2011; Naser & Concha, 2012).

Con todo esto, la idea de alcanzar “gobiernos electrónicos abiertos” ha surgido recientemente como una vía, una alternativa, un camino posible que conduzca a la construcción de mejores formas de gobierno en la era digital.



Estas iniciativas procuran ser más transparentes, confiables, participativas y colaborativas en la administración de lo público (Gobierno Vasco, 2010; Orszag, 2009). Así, se presentan a nivel internacional como un nuevo intento de llevar más apertura, diálogo y participación ciudadana a la política. En este sentido, los gobiernos abiertos parecen enfrentarse al secretismo, la clausura y, sobre todo, a la brecha que separa a los ciudadanos y a sus gobernantes dentro del sistema de representación política (Gobierno Vasco, 2010; Orszag, 2009; Ramírez-Alujas, 2010). Finalmente, la relevancia de la participación ciudadana en la era digital, ha sido motivada por la creciente y constante interacción por parte de millones de usuarios a nivel mundial, que frecuentemente ya sea de manera directa o indirecta, tienen acción y participación dentro de los medios digitales de acción política ya sea a favor o en contra del oficialismo de turno.

### *Definición de participación ciudadana y sociedad digital*

Es importante iniciar la conceptualización, con lo establecido por el órgano rector dentro del Ecuador, mismo que según el Consejo de Participación Ciudadana y Control Social (2023), la participación ciudadana es principalmente un derecho de las ciudadanas y ciudadanos, en forma individual y colectiva, de participar (ser parte de, incidir) de manera protagónica en la toma de decisiones, planificación y gestión de los asuntos públicos, y en el control popular de las

instituciones del Estado y la sociedad, así como de sus representantes, siendo un proceso permanente de construcción del poder ciudadano.

Por otro lado, la sociedad digital hace referencia a una organización social en la que sus estructuras e intercambios diarios están condicionados y conformados por sistemas automatizados, Tecnologías de la Información y Comunicación (TIC), dispositivos y otras tecnologías digitales. El proceso de digitalización implica un movimiento de adaptación e integración de tecnologías avanzadas en todas las áreas de la sociedad y la cultura. Este proceso transformador influye en el pensamiento, las decisiones y las acciones de las personas que habitan estas sociedades de la información (Educated, 2021).

De la misma manera, las sociedades digitales son el resultado de la transformación tecnológica que afecta todos los aspectos de la vida de las personas. Este cambio de paradigma, a veces llamado la Cuarta Revolución Industrial, destaca la necesidad de desarrollar nuevas estrategias sociales en las que la digitalización, la inteligencia artificial y la robótica son pilares fundamentales en todo el proceso de globalización y difusión de la información. En estos entornos cambiantes, surgen tantas ventajas como desafíos, y para comprender sus implicaciones, es necesario conocer qué son las sociedades digitales, sus características y cómo se puede garantizar su seguridad (Smowl Tech, 2024).

Tomando en cuenta lo anterior, se llega a una sola vía, conocida como plataformas digitales, y especialmente las redes sociales, que han modificado la dinámica social migrando muchas actividades cotidianas de su vida privada a internet (comunicación, educación, compras), en este sentido, Rabotnikof (2000) ejemplificó y se pudo distinguir entre lo público y lo privado, enfocado al uso de las redes sociales, mismas que desdibujan esta línea de tal forma que ahora se debe buscar una nueva frontera entre lo que le compete al Estado y lo que no, al respecto Ravel (2022) refiere que, las redes “Fungen como proveedoras de servicios privados con objetivos concretos: compartir contenidos (Facebook), crear conversaciones públicas (Twitter), generar comunidad con el público (Instagram), etcétera” (p. 8).

Finalmente, el Gobierno abierto parece encauzar a las democracias en la búsqueda de mejores procesos de gestión pública, procesos político-electorales y procesos de toma de decisiones en un contexto rodeado por la tecnología, el uso de herramientas tecnológicas se emplea para llegar a más ciudadanos y crear espacios que propicien esta relación; gobierno y sociedad. Entonces, en este contexto el principio de participación (Serrano Rodríguez, 2015), se debe trabajar para lograr una ciudadanía activa, interesada por los problemas de su entorno y dispuesta a aportar sus conocimientos y experiencias alternativas de solución, implementarlas y cocrear una democracia funcional y fuerte

en donde el ciudadano es un actor efectivamente relevante en el sistema político.

### *Relación entre democracia participativa, gobernanza digital y ciudadanía activa*

Para llevar a cabo estos tres conceptos en una misma línea de estudio, se tomará en cuenta en primer lugar a lo que hace referencia la democracia participativa, misma que constituye un modelo político que busca ampliar la intervención ciudadana en la toma de decisiones, no limitándola al acto electoral sino integrándola en distintos momentos de deliberación, control y fiscalización. Como señala Pateman (2012), la democracia participativa parte de la premisa de que la legitimidad del poder político se robustece en la medida en que las personas participan de manera directa en la construcción de lo público. Este enfoque resulta especialmente relevante en sociedades contemporáneas caracterizadas por la demanda de mayor transparencia, rendición de cuentas y apertura de los procesos estatales.

En este contexto, emerge la gobernanza digital, entendida como el uso estratégico de las tecnologías de la información y la comunicación (TIC) para transformar las relaciones entre gobierno y ciudadanía. La Organización para la Cooperación y el Desarrollo Económico (OCDE et al., 2020) sostiene que la gobernanza digital no se reduce a digitalizar

servicios, sino que implica rediseñar procesos institucionales con miras a facilitar la participación, promover la innovación democrática y asegurar la inclusión social. De este modo, la gobernanza digital se configura como el espacio institucional y tecnológico en el que se desarrollan nuevas formas de interacción política.

Por otro lado, la ciudadanía activa, refiere a la disposición y capacidad de los individuos de intervenir de manera crítica, informada y constructiva en los asuntos colectivos. En la sociedad digital, esta se materializa mediante la participación en consultas en línea, la interacción en plataformas de gobierno electrónico, la deliberación en foros digitales y el activismo a través de redes sociales (Dahlgren, 2013). La ciudadanía activa en entornos digitales no solo es receptora de información gubernamental, sino también generadora de propuestas, evaluadora de políticas y promotora de demandas colectivas.

La relación entre estos tres conceptos es, por tanto, interdependiente pero interrelacionados, en donde la gobernanza digital actúa como facilitadora tecnológica e institucional que abre canales de participación; la ciudadanía activa es el sujeto que hace uso de esos espacios para deliberar y decidir; y la democracia participativa es el marco político que otorga sentido y legitimidad a dicho proceso. Según Sæbø et al. (2008), la e-participación constituye un campo emergente que, si bien enfrenta retos como la exclusión

digital y la calidad de la deliberación, representa una oportunidad para consolidar sistemas democráticos más inclusivos y transparentes. En consecuencia, la articulación de estos tres elementos permite avanzar hacia modelos de democracia digital donde los ciudadanos no sean únicamente votantes, sino actores permanentes en la gestión pública.

Finalmente, la democracia participativa, gobernanza digital y ciudadanía activa, al apropiarse de las herramientas de la gobernanza per sé, contribuye a ampliar los espacios de participación y a fortalecer los fundamentos de una democracia participativa adaptada a los desafíos de la sociedad del conocimiento, tomando en cuenta factores sociales, políticos y económicos para el desarrollo correcto de los procesos en sí, detallando por ejemplo, la incidencia directa que se efectúa en los sistemas de gobernanza así como los preceptos sean individuales o colectivos.

## **Mecanismos de participación ciudadana en la sociedad digital**

### *Plataformas de gobierno electrónico*

En la actualidad, la participación ciudadana mediada por tecnologías digitales se ha consolidado como un componente esencial de las formas contemporáneas de gobernanza. El uso de plataformas digitales no solo permite a los ciudadanos expresar sus opiniones e ideas de manera más

directa, sino que también redefine las dinámicas de poder entre la ciudadanía y las instituciones. No obstante, es evidente que existe la falta de retroalimentación efectiva y bidireccional limita el impacto real de dicha participación sobre las decisiones políticas, debilitando su legitimidad democrática (Shin et al., 2024). En este contexto, la digitalización representa tanto un reto como una oportunidad para reducir brechas estructurales.

Por otra parte, el desarrollo de habilidades digitales en poblaciones mayores sigue siendo limitado. Zhao et al. (2023) destacan que las estrategias macro no siempre consideran los factores micro que inciden en la aceptación tecnológica, lo cual exige una mayor contextualización territorial de las políticas digitales. Asimismo, Benlahcene et al. (2024) advierten sobre las debilidades teóricas y operativas en la implementación de plataformas de e-participación. Aunque existe confianza en las instituciones, la escasa calidad del contenido digital y la falta de interacción efectiva dificultan su apropiación por parte de los ciudadanos. Además, la alfabetización digital, presenta marcadas desigualdades según nivel socioeconómico y contexto geográfico, lo cual exige intervenciones educativas específicas (Shi et al., 2024).

De esta manera, el primero a destacar es el Gobierno Electrónico en Ecuador, que debe vincularse de manera directa con elementos existentes que contemplan leyes, planes, lineamientos y normativas nacionales e

internacionales, que buscan alinear al gobierno con los distintos esfuerzos y realidad normativa existente. La base fundamental para crear un modelo que ayude a la organización de Gobierno Electrónico en Ecuador debe entender cómo se relaciona con estos cinco elementos: Constitución del Ecuador, Plan Nacional del Buen Vivir, estrategias e indicadores de la Organización de las Naciones Unidas (ONU), principios del Centro Latinoamericano de Administración para el Desarrollo (CLAD) y las definiciones de gobierno abierto, gobierno cercano, gobierno eficaz y eficiente (Gobierno Electrónico, 2023). Así, Ecuador ha avanzado significativamente en la implementación de plataformas digitales que facilitan la interacción entre el Estado y la ciudadanía. Estas herramientas buscan mejorar la eficiencia administrativa, promover la transparencia y fomentar la participación ciudadana.

Con esto, la plataforma centraliza los servicios digitales ofrecidos por el Estado, permitiendo a los ciudadanos acceder a información, realizar trámites en línea y participar en procesos de consulta pública. Es una de las iniciativas clave del Ministerio de Telecomunicaciones y de la Sociedad de la Información (MINTEL) para promover la digitalización del sector público; Por otro lado, también se toma en consideración al Portal de Gobierno Electrónico, que es el portal oficial del gobierno ecuatoriano, donde los ciudadanos pueden acceder a una amplia gama de servicios y trámites en línea, desde la obtención de documentos hasta la consulta de



servicios públicos. Esta plataforma busca simplificar los procesos administrativos y mejorar la accesibilidad de los servicios gubernamentales.

Así también, hay otra alternativa de plataforma que es denominada como ParticipaEC, lanzada en 2025, y es una plataforma digital diseñada para fortalecer la participación ciudadana. Permite a los ciudadanos involucrarse en la toma de decisiones públicas, expresar opiniones y colaborar en la creación de políticas públicas. Esta herramienta representa un avance hacia una gobernanza más inclusiva y participativa (MINTEL, 2025). Y finalmente, el Sistema Nacional de Información (SNI), es una plataforma que centraliza y organiza la información pública del país. Facilita el acceso a datos estadísticos, informes y documentos oficiales, promoviendo la transparencia y el acceso a la información pública. Es una herramienta esencial para la toma de decisiones basada en evidencia (Secretaría Nacional de la Administración Pública y Planificación [SENPLADES], s.f.).

### *Acceso a la información y transparencia digital*

Es importante tomar en cuenta que una ciudadanía informada, es una condición necesaria para un futuro “sin dejar a nadie atrás”, basado en los principios democráticos y los derechos humanos de igualdad y universalidad. Las sociedades no podrán gozar de una verdadera libertad si no cuentan con información adecuada acerca de la “cosa pública”

para tomar decisiones informadas en todos los ámbitos, sean políticos, económicos, sociales o ambientales. La transparencia y el acceso a la información, en especial la información pública, son, por lo tanto, instrumentos centrales en un Estado de derecho.

La transparencia es, a su vez, una poderosa herramienta para mejorar la eficacia y eficiencia de las políticas públicas. Cuando los procesos gubernamentales están abiertos al escrutinio público, no solo se fortalecen la confianza y la legitimidad, sino que también se crea un ciclo de retroalimentación que impulsa la mejora continua. Con acceso a información sobre decisiones, gastos y resultados, se pueden identificar áreas de ineficiencia, corregir errores, optimizar recursos y limitar el riesgo de corrupción. “La transparencia no es solo una cuestión ética, es también un motor de innovación y mejora para los gobiernos, que se ven impulsados a ser más responsables y a rendir cuentas de su actuación” (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2024, párr. 2).

De la misma manera, se han observado adopciones y modificaciones de marcos legislativos que limitan el derecho de acceso a la información. En algunos casos, esto implica proponer diseños institucionales ineficaces, como la ausencia de órganos garantes independientes que supervisen el cumplimiento de este derecho, o incluir excepciones demasiado amplias, que dejan a la discrecionalidad de las

autoridades el cumplir o no con su obligación de transparencia. En otros casos, se introducen requisitos formales engorrosos para realizar solicitudes, o se establecen sanciones a solicitantes en caso de un supuesto abuso, no claramente definido, del derecho de acceso a la información.

De esta manera, se han identificado algunas complicaciones que están vinculadas a ataques a la independencia de órganos garantes existentes, que son considerados como la espina dorsal del derecho de acceso a la información. Estos órganos son limitados en ocasiones de manera legal y, frecuentemente, de forma práctica, reduciendo la transparencia de sus procedimientos, limitando sus capacidades en recursos financieros o humanos, o bien afectando sus acciones de rendición de cuentas. Para poder monitorear la ejecución de las leyes de acceso a la información y comprender el impacto es necesaria la adecuada generación de datos y estadísticas, ya que muestran que la mayoría de los países cuenta con leyes de acceso a la información, pero que los retos de su ejecución continúan siendo importantes (UNESCO, 2024).

## **Retos y desafíos**

Uno de los principales retos para la implementación del gobierno electrónico en Ecuador es la brecha digital, que se manifiesta en la disparidad de acceso a internet y dispositivos tecnológicos entre zonas urbanas y rurales, así

como entre diferentes grupos socioeconómicos. Esta desigualdad limita la participación plena de la ciudadanía en los servicios digitales y puede profundizar la exclusión social (MINTEL, 2025). La falta de infraestructura tecnológica adecuada impide que muchos ciudadanos se beneficien de las plataformas de gobierno electrónico, afectando la equidad y la inclusión digital.

Otro desafío importante es la desinformación y la polarización, tomando en cuenta por ejemplo, las diferencias y fragmentaciones entre zonas urbanas y rurales. Aunque las tecnologías digitales facilitan la circulación de información, también amplifican la propagación de noticias falsas y discursos polarizados, lo que puede afectar la calidad de la participación ciudadana y la confianza en las instituciones públicas pero a su vez, el manejo a la conveniencia de las privadas (Macías-Ibarra & Guadarrama Chaparro, 2025). La difusión de información no verificada en redes sociales y plataformas digitales es un obstáculo para consolidar una democracia digital basada en la participación informada.

De la misma manera, la seguridad de la información y protección de datos personales constituye otro reto crítico, tomando en cuenta que, la implementación del gobierno electrónico requiere garantizar que los datos de los ciudadanos estén protegidos frente a vulneraciones y ataques cibernéticos. Según MINTEL (2025), fortalecer la ciberseguridad y establecer políticas claras de protección de

información es fundamental para la confianza ciudadana y la legitimidad de los procesos digitales, ya que por su naturaleza misma, no hay mayor control por parte de los individuos frente al Gobierno o a la Institución que lleve a cargo cualquier tipo de proceso democrático.

Asimismo, la capacitación y alfabetización digital es un desafío que afecta directamente la efectividad de los servicios electrónicos, inclusive tomando en cuenta aspectos técnicos como la calidad de acceso a internet, el uso de los dispositivos, y quien está frente a los mismos, ya que no todas las personas tienen el conocimiento de como hacerlo. Sin programas de formación adecuados, gran parte de la población no puede utilizar correctamente las plataformas digitales del Estado, limitando su participación. Por esto, la alfabetización digital no solo implica habilidades técnicas, sino también la comprensión de derechos digitales y buenas prácticas de participación en línea.

Finalmente, la confianza y transparencia en la gestión pública digital es clave para incentivar la participación ciudadana. La percepción de opacidad o de procesos poco claros puede desalentar la interacción con las plataformas gubernamentales. Por ello, se reconoce la necesidad de implementar políticas de transparencia y rendición de cuentas que fortalezcan la legitimidad de la gobernanza digital en Ecuador, sin distinción de acceso y garantizando la universalidad del mismo (Jara Iñiguez, 2023).

## *Brecha digital y exclusión en la participación política*

En primer lugar, la ambivalencia inherente a las TIC vinculada a la participación política. En efecto se constituye, como factor de cambio en la economía global basada en el conocimiento y como factor de competitividad, una oportunidad de progreso social, a la vez que un riesgo de profundización en la fragmentación y la desigual distribución de la riqueza; efecto bidireccional en el que la accesibilidad universal, el acceso a internet y la extensión de la infraestructuras de la información y la comunicación juegan un papel determinante en este sentido. Es innegable que las TIC han generado expectativas de mejora social, al posibilitar, en términos teóricos, el acceso universal a la información, fortalecer la sociedad civil y la democracia, avanzar en la inclusión social, optimizar recursos, brindar nuevas oportunidades culturales, generar nuevos empleos y más progreso social, pero también es posible que genere efectos contrarios al agrandar las diferencias entre países, regiones y grupos sociales (Olarte Encabo, 2017).

Según UNESCO (2013), en su informe Universalidad de Internet: un medio para crear sociedades del conocimiento y la agenda de desarrollo sostenible después de 2015, centra la importancia del cambio tecnológico en Internet, al que define como un sistema global de dispositivos interconectados que ofrece servicios a los usuarios y usuarias de todo el mundo. Fenómeno que trata describirse de forma

aséptica, señalando cómo después de tres décadas de evolución y adopción, de la participación de usuarios en la política digital se ha convertido en una red de redes diversificada y en evolución que consta de millones de conexiones privadas, públicas, académicas y de administraciones, de un alcance que va desde local a mundial, y que se gestiona a través de una amplia gama de tecnologías.

En particular, UNESCO (2013, p. 13), hace referencia a un informe centrado en la descripción del fenómeno del acceso a las sociedades digitales a informarse en los procesos de democracia activa, participación política y educación colectiva de transparencia y acceso a la información pública, todo esto, incide igualmente en los retos y riesgos de fractura a los diferentes estratos sociales y geográficos en la sociedad global, estableciendo el principio de universalidad de internet y la accesibilidad para todas las personas y todas en todas las dimensiones. Concretamente afirma la necesidad de que la información sea “accesible para todos” como parte de la “universalidad de Internet”, lo que implica acceso abierto, como vía de evitar discriminaciones digitales, y “que los modelos sostenibles y fiables de negocio puedan financiar el acceso universal y además puedan garantizar la accesibilidad a través del mantenimiento de una gama diversa de contenidos y de servicios” (UNESCO, 2013, p. 15).

Asimismo, se destaca la importancia necesaria de la implicación pública y de recursos públicos “que va desde las

asociaciones entre el sector público y el privado hasta la producción entre iguales basada en el dominio, y una normativa y unas medidas de política adecuadas, son necesarias para una universalidad eficaz”. De este modo, las distintas organizaciones internacionales coinciden en señalar que las TIC por sí solas no constituyen un remedio para las grandes fracturas sociales, sino que, al contrario, pueden incidir en un agravamiento de las brechas sociales preexistentes. Y el reto, es crear un marco jurídico e institucional que garantice, entre otros principios, la no discriminación tecnológica, porque las TIC también tienen un potencial importante en orden a aumentar las oportunidades de todos y mejorar el nivel de vida de forma generalizada (MINTEL, 2025; UNESCO, 2013, 2024).

### *Derechos digitales y protección de datos*

Es importante tomar en consideración la responsabilidad de aplicación de los derecho digitales en cualquier proceso que involucre el sistema de protección de datos. Tomando como prioridad por ejemplo: Derecho al olvido, libertad de expresión digital, protección de datos. En la época actual, caracterizada por la conectividad a la red, es necesario que las leyes ecuatorianas y de todos los países se adapten para proteger los derechos digitales de las personas. Esta clase de garantías son esencialmente una extensión de las recogidas dentro Declaración Universal de los Derechos Humanos de la Organización de las Naciones Unidas,



adaptadas a la era de internet. A pesar de su importancia, no existe un convenio internacional que recoja estos derechos en un documento único. Las tecnologías de la información y la comunicación han cambiado la forma en que se vulneran derechos fundamentales como la libertad de expresión, el acceso a la información o el derecho a la privacidad, dejando a las personas más expuestas (UNIR, 2022). Por ello, se requiere garantizar los derechos digitales dentro de la participación ciudadana en la era digital. Aunque no existe una carta única de derechos digitales, en líneas generales son:

En primer lugar, el acceso universal e igualitario a internet. De acuerdo con el Instituto Nacional de Estadísticas y Censos INEC (2020), los hogares en Ecuador tuvieron una conexión a internet: 61,7% en el área urbana y 34,7% en el área rural. Aunque es un gran salto adelante, esto no es suficiente, ya que el uso de la tecnología viene aparejado de importantes cambios sociales, culturales, económicos y profesionales.

Las personas necesitan navegar por la web sin importar sus ingresos, ubicación. Sin un acceso asequible y sencillo de utilizar, no se pueden disfrutar plenamente de ciertos derechos en Ecuador, como la libertad de opinión, por lo que es esencial abatir la brecha digital. (UNIR, 2022, párr. 4-5)

Por otro lado, y muy importante, la libertad de expresión digital, de manera general, la libertad de expresión está considerada como esencial para el desarrollo y la protección de los derechos humanos. Cuando se vetan páginas, contenidos o el uso de redes sociales, se incurren en violaciones de los derechos de comunicación, libertad de expresión e información y libre asociación. El último informe elaborado por la Relatoría Especial para la Libertad de Expresión (RELE) de la Comisión Interamericana de Derechos Humanos (CIDH) expone que Ecuador se posicionó en 2021 como uno de los 18 países que mejoró en los índices de libertad de expresión a través de Internet, de acuerdo con la calificación de Freedom House. Cabe destacar que fue el país que demostró el avance más significativo (CIDH, 2022).

Así también, se toma en consideración al Derecho al anonimato, privacidad y protección de datos. Para realizar transacciones fiables y seguras en línea, hay que cifrar ciertas comunicaciones, por lo que se vulnera este derecho. Entre los derechos de un ciudadano digital está el que los usuarios puedan controlar quién almacena sus datos personales y poder eliminarlos cuando lo desee. En el país, se promulgó la Ley orgánica de protección de datos personales, que busca proteger a la población. Finalmente, el derecho al olvido, es trascendental que la gente tenga la posibilidad de borrar su información privada de los buscadores, bases de datos y directorios, ya que esto está vinculado con su derecho a la

intimidación. Este principio está reconocido en la ya mencionada Ley de Protección de Datos Personales (Ecuador, 2021).

## **Conclusiones y perspectivas**

La sociedad digital ha transformado profundamente los mecanismos de participación ciudadana, generando nuevas oportunidades para la deliberación, el control social y la incidencia en la toma de decisiones públicas. Las tecnologías de la información y la comunicación (TIC) han permitido ampliar los espacios de interacción entre ciudadanía y Estado, facilitando procesos más inclusivos, ágiles y transparentes. Sin embargo, esta expansión digital no ha estado exenta de desafíos: la brecha tecnológica, la falta de alfabetización digital y los riesgos asociados a la desinformación política continúan limitando el ejercicio pleno de la participación democrática en entornos digitales.

De esta manera, los mecanismos tradicionales como las consultas populares, cabildos o presupuestos participativos han encontrado en las plataformas digitales un canal complementario para fortalecer la gobernanza democrática. No obstante, la efectividad de estos instrumentos depende de su regulación jurídica, de la voluntad política de las instituciones y de la capacidad de la ciudadanía para apropiarse de los medios digitales como herramientas de poder colectivo. La participación en línea no puede sustituir los espacios presenciales de deliberación, pero

sí puede potenciarlos al favorecer la transparencia, la trazabilidad y el acceso a la información pública.

Así también, los mecanismos de participación ciudadana en la sociedad digital representan una oportunidad para consolidar democracias más horizontales y deliberativas, siempre que se acompañen de políticas de inclusión tecnológica, educación cívica digital y marcos normativos que garanticen derechos digitales como la privacidad, la libertad de expresión y el acceso universal a la información. En el futuro inmediato, los procesos de participación ciudadana tenderán a evolucionar hacia modelos híbridos que integren herramientas digitales y presenciales, buscando equilibrar eficiencia tecnológica con legitimidad social. La inteligencia artificial y las plataformas de gobierno abierto podrían convertirse en instrumentos clave para garantizar procesos más transparentes, verificables y participativos.

Finalmente, será necesario repensar el rol del Estado y de las instituciones públicas en la mediación de la participación digital, fortaleciendo los marcos de gobernanza y ética tecnológica. Las universidades, organizaciones sociales y actores de la sociedad civil deberán asumir un papel protagónico en la promoción de una ciudadanía digital crítica y activa, capaz de incidir en la agenda pública más allá del voto o la opinión en redes sociales. Finalmente, el reto central consistirá en democratizar el acceso a las tecnologías, reducir las brechas digitales y asegurar que la participación

ciudadana en la era digital no reproduzca las desigualdades existentes, sino que contribuya efectivamente a la construcción de una democracia más justa, inclusiva y participativa.

## Referencias

- Banda Castro, A. L., González Beltrones, A. V., Valenzuela Miranda, G. A., Morales Zamorano, M. A., & Avendaño Enciso, A. (2010). Participación ciudadana: Perspectiva cuantitativa de estudio. *Enseñanza e Investigación en Psicología*, 15(2), 377–392.
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.
- Benlahcene, A., Awang, H., Mansor, N. S., Nadzir, M. M., Yamin, F. M., & Haruna, I. U. (2024). Citizens' e-participation through e-government services: A systematic literature review. *Cogent Social Sciences*, 10(1).  
<https://doi.org/10.1080/23311886.2024.2415526>
- Berners-Lee, T. (2000). *Weaving the Web: The original design and ultimate destiny of the World Wide Web*. HarperCollins.
- Castells, M. (2001). *La galaxia Internet: Reflexiones sobre Internet, empresa y sociedad*. Areté / Plaza & Janés.
- Comisión Interamericana de Derechos Humanos. (2022, 26 de mayo). *Informe Anual 2021: Informe de la*

*Relatoría Especial para la Libertad de Expresión*  
(OEA/Ser.L/V/II. Doc. 64 rev.1). Organización de los  
Estados Americanos.  
[https://www.oas.org/es/cidh/expresion/informes/IA  
2021ESP.pdf](https://www.oas.org/es/cidh/expresion/informes/IA2021ESP.pdf)

Consejo de Participación Ciudadana y Control Social. (2023, julio). *Glosario Participación Ciudadana y Control Social*. <https://www.cpccs.gob.ec/?mod=glosario>

Dahlgren, P. (2013). *The political web: Media, participation and alternative democracy*. Palgrave Macmillan.

Delfino, G. (2013). *Participación política y factores psicosociales: Un estudio con estudiantes universitarios*. Editorial Académica Española.  
<https://hdl.handle.net/11336/116720>

Delfino, G., & Zubieta, E. (2010). Participación política: Concepto y modalidades. *Anuario de Investigaciones*, 17, 211–220.  
[https://www.scielo.org.ar/scielo.php?script=sci\\_artte  
xt&pid=S1851-16862010000100020](https://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-16862010000100020)

Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales* (Registro Oficial Suplemento 459, 26 de mayo de 2021). Ministerio de Telecomunicaciones y de la Sociedad de la Información.  
[https://www.telecomunicaciones.gob.ec/wp-  
content/uploads/2021/06/Ley-Organica-de-Datos-  
Personales.pdf](https://www.telecomunicaciones.gob.ec/wp-content/uploads/2021/06/Ley-Organica-de-Datos-Personales.pdf)

- Educated. (2021, agosto). *Sociedades digitales*.  
<https://elexploradordevisita.weebly.com/sociedades-digitales.html>
- Gobierno Electrónico. (2023). *Gobierno Electrónico en Ecuador*.  
<https://www.gobiernoelectronico.gob.ec/gobierno-electronico-en-ecuador/>
- Gobierno Vasco. (2010). *¿Qué es Irekia?*  
<https://www.irekia.euskadi.eus/es/site/page/about>
- Jara Iñiguez, I. (2023). Estado de la transparencia y acceso a la información pública en Ecuador pos-pandemia. *GIGAPP Estudios Working Papers*, 10(267–272), 396–415.  
<https://www.gigapp.org/ewp/index.php/GIGAPP-EWP/article/view/330>
- Jorquera Beas, D. (2011). *La re-invencción de la política: Internet y la nueva esfera pública*. Editorial Planeta.
- Lessig, L. (2001). *The future of ideas: The fate of the commons in a connected world*. Random House.
- Lessig, L. (2008). *Remix: Making art and commerce thrive in the hybrid economy*. Penguin Press.
- Macías-Ibarra, G. A., & Guadarrama Chaparro, A. (2025). La participación ciudadana y el gobierno abierto en la era digital: Una mirada desde la IA. *Universitas-XXI, Revista de Ciencias Sociales y Humanas*, (43), 123–150. <https://doi.org/10.17163/uni.n43.2025.05>

- Ministerio de Telecomunicaciones y de la Sociedad de la Información. (2025, 16 de abril). *Ecuador fortalece la participación ciudadana con la nueva plataforma digital ParticipaEC* [Comunicado de prensa N.º 081]. Gobierno Electrónico Ecuador.  
<https://www.telecomunicaciones.gob.ec/ecuador-fortalece-la-participacion-ciudadana-con-la-nueva-plataforma-digital-participaec>
- Naser, A., & Concha, G. (2012). Panorama de gobierno electrónico en la región: Resultados e impactos. En G. Concha & A. Naser (Eds.), *El desafío hacia el gobierno abierto en la hora de la igualdad* (pp. 1–30). Naciones Unidas.  
<http://www.cepal.org/ddpe/publicaciones/xml/9/46119/W465.pdf>
- OCDE, Naciones Unidas, CAF, & Unión Europea. (2020). *Perspectivas económicas de América Latina 2020: Transformación digital para una mejor reconstrucción*. OECD Publishing.  
<https://doi.org/10.1787/f2fdced2-es>
- Olarte Encabo, S. (2017). Brecha digital, pobreza y exclusión social. *Temas Laborales: Revista Andaluza de Trabajo y Bienestar Social*, (138), 285–313.  
<https://www.juntadeandalucia.es/sites/default/files/inline-files/2023/06/09.%20Brecha%20digital%2C%20pobreza%20y%20exclusi%20n%20social%20.pdf>



- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO]. (2013). *Universalidad de Internet: Un medio para crear sociedades del conocimiento y la agenda de desarrollo sostenible después de 2015*. [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/internet\\_universality\\_es\\_01.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/internet_universality_es_01.pdf)
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO]. (2024, 28 de septiembre). *La transparencia y el acceso a la información pública en América Latina y el Caribe: Pilares de la democracia*. <https://www.unesco.org/es/articles/la-transparencia-y-el-acceso-la-informacion-publica-en-america-latina-y-el-caribe-pilares-de-la>
- Orszag, P. (2009). *Open government directive* (OMB Memorandum M-10-06). Office of Management and Budget. [https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/memoranda\\_2010/m10-06.pdf](https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf)
- Pateman, C. (2012). Participatory democracy revisited. *Perspectives on Politics*, 10(1), 7–19. <https://doi.org/10.1017/S1537592711004877>
- Rabotnikof, N. (2000). Público-privado. En L. Baca Olamendi, J. Bokser-Liwerant, F. Castañeda, I. H. Cisneros, & G. Pérez Fernández del Castillo (Comps.),

*Léxico de la política* (pp. 604–607). Fondo de Cultura Económica.

- Ramírez-Alujas, Á. (2010). Innovación en la gestión pública y open government (gobierno abierto): Una vieja nueva idea. *Buen Gobierno*, (9), 1–19. [https://www.gigapp.org/administrator/components/com\\_jresearch/files/publications/Gobierno%20Abierto%20RBG%202010%20\[Original\].pdf](https://www.gigapp.org/administrator/components/com_jresearch/files/publications/Gobierno%20Abierto%20RBG%202010%20[Original].pdf)
- Ravel, D. (2022). Presentación. En *Democracia y redes sociales* (pp. 7–12). Instituto Nacional Electoral.
- Sæbø, Ø., Rose, J., & Flak, L. S. (2008). The shape of eParticipation: Characterizing an emerging research area. *Government Information Quarterly*, 25(3), 400–428. <https://doi.org/10.1016/j.giq.2007.04.007>
- Secretaría Nacional de la Administración Pública y Planificación. (s. f.). *SNI: Información del Ecuador al alcance de todos*. <https://www.planificacion.gob.ec/sni-informacion-del-ecuador-al-alcance-de-todos>
- Serrano Rodríguez, A. (2015). La participación ciudadana en México. *Estudios Políticos*, (34), 93–116. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-16162015000100005](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-16162015000100005)
- Shi, Z., Du, X., Li, J., Hou, R., Sun, J., & Marohabutr, T. (2024). Factors influencing digital health literacy among older adults: A scoping review. *Frontiers in*

*Public Health*, 12, 1–14.

<https://doi.org/10.3389/fpubh.2024.1447747>

Shin, B., Floch, J., Rask, M., Bæck, P., Edgar, C., Berditchevskaia, A., Mesure, P., & Branlat, M. (2024). A systematic analysis of digital tools for citizen participation. *Government Information Quarterly*, 41(3), Article 101954.

<https://doi.org/10.1016/j.giq.2024.101954>

Smowl Tech. (2024, 24 de febrero). *Sociedades digitales: Qué son, características y cómo garantizar su seguridad*.

<https://smowl.net/es/blog/sociedades-digitales/>

UNIR. (2022, 19 de septiembre). *Los derechos digitales en Ecuador: ¿Cuáles son y cómo se regulan?*

<https://ecuador.unir.net/actualidad-unir/derechos-digitales/>

van Deth, J. (2001, 6–11 de abril). *Studying political participation: Towards a theory of everything?* [Trabajo presentado]. Joint Session of Workshops of the European Consortium for Political Research, Grenoble, Francia.

Vercelli, A. (2007). La participación política ciudadana en la era digital. *Revista Virtualis*.

<https://www.revistavirtualis.mx/index.php/virtualis/article/view/72/137>

Vercelli, A. H. (2009). *Repensando los bienes intelectuales comunes: Análisis sociotécnico sobre el proceso de coconstrucción entre las regulaciones de derecho de*

*autor y derecho de copia y las tecnologías digitales para su gestión.* <http://hdl.handle.net/11336/194758>

Zhao, T., Ma, Y., Ding, J., Yang, X., & Sheng, D. (2023). A survey on the digital skill status of rural elderly groups under the digital village strategy: An example from Sichuan, Gansu and Zhejiang provinces. *BCP Business & Management*, 42, 28–32. <https://doi.org/10.54691/bcpbm.v42i.4557>

# **Emociones y movilización política en el sexenio de AMLO (2018-2024). La reconfiguración de la política mexicana en el contexto de la Cuarta Transformación**

*Emotions and Political Mobilization during AMLO's Administration (2018–2024). The Reconfiguration of Mexican Politics in the Context of the Fourth Transformation*

**Ekhiñe Graell Larreta<sup>46</sup>**

## **Resumen**

Este capítulo analiza el papel central que desempeñaron las emociones en la configuración política del

---

<sup>46</sup> UAM-X. <https://orcid.org/0009-0003-6831-396X>

sexenio de Andrés Manuel López Obrador (2018–2024). A partir de un marco teórico que aborda la relación entre subjetividad, cultura política y afectos, se examina cómo la Cuarta Transformación utilizó narrativas emocionales para articular identidades políticas, movilizar apoyos y construir un horizonte moral diferenciado entre “pueblo bueno” y adversarios. Se describen las transformaciones discursivas de las campañas de 2006, 2012 y 2018, enfatizando elementos como la esperanza, el amor, la reivindicación ética y la simbología religiosa. Asimismo, se muestra cómo la comunicación presidencial —particularmente las conferencias matutinas— reforzó la polarización afectiva y desplazó debates programáticos hacia significados morales y percepciones subjetivas. El análisis permite comprender la emocionalización como una estrategia de poder y como un proceso estructural de la política contemporánea en México.

**Palabras clave:** Emocionalización de la política; emociones en la política; movilización política; sexenio de AMLO; Cuarta Transformación.

## **Abstract**

This chapter analyzes the central role that emotions played in shaping political dynamics during Andrés Manuel López Obrador’s administration (2018–2024). Drawing on a theoretical framework that links subjectivity, political culture, and affect, it examines how the Fourth Transformation

employed emotional narratives to articulate political identities, mobilize support, and construct a moral horizon dividing the “good people” from political adversaries. The chapter traces the discursive evolution across the 2006, 2012, and 2018 presidential campaigns, highlighting elements such as hope, love, ethical renewal, and religious symbolism. It also shows how presidential communication—particularly the daily morning press conferences—reinforced affective polarization and shifted public debate from programmatic discussions toward moral meanings and subjective perceptions. This analysis demonstrates how emotionalization operated both as a political strategy and as a structural process within contemporary Mexican politics.

**Keywords:** The emotionalization of politics; emotions in politics; political mobilization; AMLO’s six-year term; the fourth transformation.

## **Introducción**

En la sociedad contemporánea se está dando un contexto político particular que se ha traducido en una desafección de los ciudadanos hacia la democracia. Autores como Ludolfo Paramio (2002) o Luigi Ferrajoli (2014), entre otros, atribuyen esta insatisfacción a causas económicas y, en particular, a la globalización. Es indudable que los sistemas democráticos se han visto afectados por la globalización dominada por los grandes centros del poder económico y que

ésta ha puesto en crisis el poder de decisión de los ciudadanos, puesto que ya no son estos últimos los que toman las decisiones políticas, sino los mismos actores económicos. En este sentido, Chantal Mouffe (1999) o Slavoj Žižek (2010) plantean que, en la política contemporánea, se han desdibujado las diferencias ideológicas entre la izquierda y la derecha, lo que ha derivado en el comienzo de una nueva era llamada postpolítica, en la que no hay existencia de ideologías antagónicas y, por ende, no hay proyectos políticos diferenciados entre los que la ciudadanía puede optar. En la misma línea, Jacques Rancière (1996) propuso que la política deja de existir cuando la sociedad se homogeniza, pues “la política existe mientras haya formas de subjetivación singulares que renueven las formas de la inscripción de la identidad entre el todo y la comunidad y la nada que la separa de sí misma” (p. 153).

Colin Crouch (2004, 2020) ha denominado a esta situación posdemocracia, un contexto político en el que las élites económicas alineadas con los gobiernos electos establecen la agenda política dirigida al enriquecimiento de esa élite. De esta manera, aunque continúa la existencia de las elecciones, en realidad, el debate electoral está reducido a los intereses de un pequeño grupo y, por ende, limitado a las cuestiones que ellos seleccionan, por lo que las contiendas electorales se convierten en una simulación de la democracia. Así, el debate público se reduce y se vacía de alternativas reales, lo que favorece la aparición de liderazgos disruptivos y



formas de movilización política no necesariamente vinculadas a programas, sino a afectos y percepciones. En este contexto político particular, hemos podido observar el auge de la extrema derecha en diversas partes del mundo o la llegada al poder de personajes políticos tan controvertidos como el estadounidense Donald Trump o el argentino Javier Milei, cuya dimensión más relevante no es lo programático, sino emocional. En suma, la política contemporánea incorpora cada vez más estrategias de identificación y movilización basadas en las emociones, tendencia que en México se manifestó de manera especialmente clara durante el sexenio de Andrés Manuel López Obrador (2018–2024).

Este texto analiza la centralidad que han adquirido las emociones en la esfera política contemporánea, y cómo estas han reconfigurado el panorama político mexicano durante el sexenio de Andrés Manuel López Obrador (2018–2024). En particular, se examina el papel que desempeñaron las subjetividades políticas, esto es, los modos en los que los individuos y colectivos se perciben, sienten y se relacionan con el poder, en la construcción y la consolidación de la llamada Cuarta Transformación<sup>47</sup>. Más que un simple

---

<sup>47</sup> La Cuarta Transformación o 4T hace referencia al proyecto de nación y visión de gobierno del expresidente Andrés Manuel López Obrador (AMLO) -fundador del partido político Movimiento Regeneración Nacional (Morena)- y de la actual presidenta de la República mexicana, Claudia Sheinbaum Pardo, del mismo partido. Este proyecto de nación establece el triunfo de AMLO en las elecciones de 2018 y, por tanto, la llegada al poder de Morena como un nuevo momento clave en la historia

proyecto de gobierno, la 4T se presentó durante el sexenio como una ruptura histórica cargada de significados afectivos: esperanza, redención moral, resentimiento acumulado y promesas de restauración comunitaria. Esta dimensión emocional es fundamental para comprender por qué el triunfo de López Obrador fue interpretado por amplios sectores como un parteaguas comparable a episodios fundacionales de México como la Independencia, la Reforma o la Revolución mexicana.

El contexto de la Cuarta Transformación ha estado marcado por una polarización afectiva sin precedentes, visible desde los primeros meses del sexenio de Andrés Manuel y cuyos efectos se prolongaron más allá del sexenio: “los electorados partidistas están altamente polarizados en el plano ideológico en el país, independientemente del tamaño que representa cada uno” (Moreno, 2024, p. 231). La figura de López Obrador generó emociones situadas en extremos opuestos: adhesiones casi reverenciales y rechazos profundos que en ocasiones derivaron en odio político (Ahuactzin & González, 2019). Esta polarización no solo estructuró el debate público, sino que permeó la vida cotidiana, configurando identidades, conversaciones y percepciones

---

de México, equiparable a algunos de los acontecimientos más importantes y transformadores de la historia del país: la Independencia, la Reforma y la Revolución mexicana. De manera que, la Cuarta Transformación, significaría el comienzo de una nueva etapa histórica y política en México, iniciada desde el 2018.

sobre la legitimidad del gobierno y sus opositores. La expansión de estas pasiones demuestra que las emociones fueron un componente central de la 4T que funcionaron como mecanismos de movilización política y de construcción de subjetividades colectivas.

El planteamiento central que guía este ensayo es que, además del indudable hartazgo de la sociedad mexicana ante los resultados obtenidos por los últimos gobiernos panistas y priistas anteriores, en particular, los de Felipe Calderón (2006-2012) y Enrique Peña Nieto (2012-2018), respectivamente, en materia de violencia, corrupción y economía, entre otras cuestiones, el énfasis que Andrés Manuel otorgó a las emociones en su discurso político durante la campaña para las elecciones de 2018, contribuyó a la movilización de amplios sectores del electorado. Más que explicar de manera causal el resultado electoral, este trabajo propone interpretar la dimensión emocional del discurso político como un mecanismo de movilización electoral que permitió construir un horizonte moral compartido entre el líder político y sus bases sociales. Lo anterior, debido a que las emociones y los sentimientos juegan un papel central en la configuración política de la sociedad contemporánea. Estudios recientes sobre comportamiento electoral confirman que, en contextos de incertidumbre, los votantes reaccionan más a percepciones y significados afectivos que a información objetiva (Panagopoulos & Wang, 2025), proceso que la 4T reforzó posteriormente a través de dispositivos como las

conferencias matutinas, la narrativa del “pueblo bueno” y la resignificación moral del adversario.

Para lograr lo anterior, el texto se divide en tres partes: en la primera de ellas, se realiza una reflexión teórica sobre la relación que existe entre la política y la emoción en la sociedad contemporánea; en la segunda parte, se estudia la influencia que tuvieron las emociones en la campaña electoral de Andrés Manuel López Obrador para los distintos comicios en los que participó, con énfasis en la del 2018, así como la forma de gobernar durante el comienzo de la denominada Cuarta Transformación. Finalmente, se ofrece una conclusión que sintetiza los hallazgos y plantea preguntas para futuras investigaciones sobre la emocionalización de la política en México y, en general, en las democracias deliberativas occidentales.

## **La política y la emoción en la sociedad contemporánea**

La política está fuertemente influenciada por la subjetividad, tanto individual como colectiva -la intersubjetividad-. Las creencias, las imágenes y las emociones tienen una gran injerencia en las decisiones políticas de los seres humanos y, por tanto, ejercen un fuerte influjo en los procesos políticos, incluyendo la esfera de la política formal o institucional. Como plantea Bolívar Echeverría (2011), lo político permea en la vida cotidiana -

tanto como la actividad política misma, como de forma imaginaria, en forma de ruptura de la cotidianeidad-; por tanto, la visión simplista de la política reducida al ámbito de las instituciones que gestionan los asuntos públicos no permite entender enteramente lo político. El reduccionismo sobre la concepción de lo político considera a la cuestión política a aquellas actividades públicas y racionales<sup>48</sup>, sin embargo:

una aproximación crítica a la cultura política no puede dejar de insistir en que la realización de lo político por la vía de la actividad especialmente política tiene necesariamente que ver, sin excepción, con [...] instituciones que pertenecen a órdenes muy diferentes [instituciones de parentesco, civiles, religiosas, laborales, etc.]. (Echeverría, 2011, p. 172)

De la misma manera, la acotación y restricción que se le ha conferido al ámbito político, como ocurre con el modelo sistémico eastoniano, minimiza los alcances de la configuración sociocultural que son vitales en algunos contextos.

---

<sup>48</sup> En este sentido, Bolívar Echeverría (2011) menciona que los estudiosos de la política moderna como Hegel han diferenciado entre la “sociedad civil” y el “estado”, separando en dos planos la vida en sociedad, una plenamente civil o burguesa y la otra plenamente política o ciudadana, otorgándole exclusividad al ámbito político.

En este sentido, el sujeto ha sido poco estudiado en el análisis político, ya que la reflexión y la teorización política se ha concentrado, principalmente, en los colectivos, y no en las personas. Sin embargo, es necesario conceder a los sujetos la dimensión que les corresponde, pues son los ciudadanos los que forman la parte fundamental de la vida política: a la vez fungen como los actores y los beneficiarios de ésta. Es necesario, entonces, considerar a las personas o a los ciudadanos como sujetos pensantes que, a su vez, poseen sentimientos y que se conducen con base en sus pensamientos y emociones: “Para el caso del ámbito político en la vida de un individuo, se trata de reconocer que precisamente su comportamiento político está determinado -entre otros factores de orden histórico, estructural y social- por su subjetividad” (Rocha, 2005, p. 99). Por ende, es necesario que en el análisis político se tomen en cuenta los aspectos socioculturales de los sujetos, es decir, de los ciudadanos, puesto que son elementos que interfieren en la política formal.

Desde finales del siglo XX se han ido desdibujando los límites entre lo público y lo privado, por lo que se podría decir que hemos regresado al “tiempo de las emociones, en el sentido psicológico e histórico del término” (Magaña, 2016, p. 14). Las instituciones que en otras épocas resultaban importantes para una identidad colectiva como la escuela, la familia, las empresas, e incluso la política misma, ya no lo son desde hace algunas décadas. Es por ello por lo que resulta más

que nunca necesario estudiar la sociedad tomando en cuenta al ser humano en su plenitud, considerando que éste es un ser tanto racional como emocional, y sobre la que se debe poner especial énfasis en la cultura política como modeladora de la política formal o institucional. Como planteó Norbert Lechner (1988), “la acción política, como toda acción humana, no obedece sólo a una lógica cognitiva sino igualmente a una lógica afectiva y simbólica. Probablemente estas dimensiones condicionen (mucho más que la razón cognitiva) nuestras experiencias diarias con el tiempo político” (p. 89). Este giro hacia lo emocional resulta especialmente visible en contextos como el mexicano, donde la subjetividad y la percepción han adquirido un peso determinante en la configuración de la cultura política reciente.

Las particularidades de las sociedades posmodernas, en las que los imaginarios colectivos tienen mayor peso que las sociedades decimonónicas y de principios del siglo XX (Maffesoli, 2013), hacen necesario analizar esa interrelación entre las subjetividades políticas y las decisiones que se toman en torno a las instituciones. Resulta relevante que en la colectividad actual las personas se agrupen alrededor de las identidades primarias como la etnia, la cultura o el territorio, y se hayan convertido en la principal fuente de significación social (Cantero, 2018). Y no sólo en la posmodernidad, ya que ha sido el propio pensamiento simbólico del ser humano el que permitió que éste se convirtiera en un “animal político”. Nuestra realidad, entonces, tiene como pilar el entramado

simbólico, que nos permite comprender e interpretar la realidad. Por ende, los sistemas simbólicos son las bases de la cultura y los que permiten la existencia de la actividad social y política, de manera que coadyuvan la construcción de las realidades políticas y estructuran la forma de comprender el mundo. Cabe destacar que la base de esos sistemas simbólicos es la comunicación (Amador, 2014).

En este sentido, para Carlos Cantero (2018), el individuo contemporáneo es:

un «Homo Emóticus», un individuo que es y hace lo que siente, lo que son sus deseos, aquello que compromete sus emociones y sentidos, principalmente desde la configuración de su mapa mental por la instantánea, ubicua, obsolescente, efímera, pero, determinante presencia de los medios de comunicación masiva en interacción con su cultura endógena. (Cantero, 2018, p. 41)

En la sociedad contemporánea, las decisiones que se suelen tomar en la vida diaria están, en buena medida, basadas más en las emociones y subjetividades que en el pensamiento enteramente racional y objetivo; las decisiones más importantes de la vida humana son tomadas con base en los sentimientos, aunque indudablemente siempre existe algo de racionalidad en ellas. De esta manera, la cultura política es el pilar de la conformación de la política formal



contemporánea, que se puede concebir como un aglomerado de emociones, creencias, imágenes, valores y discursos que los ciudadanos generan sobre la política formal.

¿Pero cómo se conforman las emociones? ¿Qué componentes son los que constituyen la respuesta afectiva? Rafael Bisquerra (2017) distingue tres elementos: el neurofisiológico, que se basa en las respuestas corporales como pueden ser la taquicardia, las exudaciones hormonales, la hipertensión, las transformaciones en el tono muscular, etc.; el comportamental, que tiene que ver precisamente con la expresión afectiva que puede tener un sujeto particular, es decir, con la conducta que tiene una persona ante una emoción concreta, su lenguaje no verbal ante ella; y, por último, el componente cognitivo, que se da cuando el sujeto adquiere cognición de la emoción experimentada y puede nombrarla y describirla a partir del lenguaje oral o escrito. Por su parte, las emociones están estrechamente relacionadas con las acciones: la experiencia emocional es, en gran medida, una acción experimentada, un estado comprobado de preparación para la acción (Frijda, 2012). Lo anterior no significa que dicha acción tenga que ocurrir necesariamente, pero sí que las emociones inducen a la acción. Que exista o no esa respuesta de acción a determinadas emociones tiene que ver, en gran medida, con el contexto en el que ocurren esas impresiones.

De esta manera, la gestión emocional o de las emociones se vuelve una forma de poder que tiene mucha

injerencia en la sociedad contemporánea. En este sentido, la manipulación emocional es una poderosa herramienta de dominación (Cantero, 2018); de ahí la importancia de conocer y entender cuál es la relación entre los sentidos y las emociones con la política formal, ya que en la sociedad posmoderna el poder radica en la capacidad de: “configurar, en un caso, o desfigurar, en otros, los mapas mentales de los actores sociales; gestionar la subjetividad de las personas, particularmente sus necesidades y deseos, enfatizando apegos emocionales o materiales, utilizando al propósito los medios de comunicación masiva” (Cantero, 2018, p. 43). Así, en el mundo contemporáneo los medios de comunicación y la comunicación en sí misma -independientemente de que ésta provenga o no de los medios de comunicación masiva-, configuran y han conformado imaginarios con base en una narrativa que apela a la dimensión afectiva.

Dado a que la concepción de nuestro mundo está construida con base en el lenguaje y en la comunicación, el poder es sinónimo de tener la capacidad de construir realidades con base en ellos, esto es, es proporcional a la capacidad de crear significados y sentidos: los mapas mentales que son compartidos por la sociedad. En este sentido, Sara Ahmed (2015) ha planteado que las emociones son experiencias culturales en lugar de estadios psicológicos, pues están socialmente constituidas por medio de círculos afectivos. Según la autora, las emociones funcionan a través de signos y cuerpos que se materializan en las superficies y los

límites que la sociedad vive como mundos. Es por ello por lo que, a menudo, las emociones, que principalmente son generadas por discursos<sup>49</sup>, constituyen un instrumento: se utilizan para persuadir o seducir a otros hacia falsas creencias o suposiciones.

La sociedad contemporánea se encuentra hiperconectada y cuenta con una accesibilidad a la información nunca vista. Sin embargo, esa saturación de información también conlleva una mayor desinformación debido al auge de las llamadas *fake news*, así como una comunicación centrada en incidir en nuestros sentimientos, dirigido a nuestro corazón. Este órgano de nuestro cuerpo, que es, a su vez, el símbolo de las emociones y los sentimientos, lo encontramos de forma frecuente a nuestro alrededor, sobre todo en los comerciales y los anuncios, pero inclusive también lo podemos hallar en los logotipos y emblemas de algunos partidos y movimientos políticos que buscan crear identificación inmediata, como se observó en campañas recientes en América Latina y Europa. Esta tendencia se reproduce también en México, donde la comunicación política incorpora estrategias de

---

<sup>49</sup> El discurso entendido como un “universo entero de entidades implicadas en un tema” (Wagner et al., 2011, p. 171) que incluye los diálogos, las teorías y las aseveraciones de un grupo social. Se trata de un discurso que alimenta los imaginarios sociales y determina el pensamiento y el simbolismo de una sociedad en concreto.

simplificación, impacto emocional e individualización<sup>50</sup>. La política, en la actualidad, está condicionada a la manipulación emocional; el mensaje político actual debe tener tres características: simplificación, impacto emocional e individualización para que, al igual que como ocurre con el lenguaje audiovisual y publicitario, sea consumido de forma más rápida y eficaz por los ciudadanos (Aira, 2020).

Así, las emociones están estrechamente relacionadas con la política formal y, hoy en día, más que nunca, se está configurando una cultura política centrada en las emociones. Las investigaciones en torno al comportamiento electoral han sugerido que en los votantes tienen más influencia las percepciones que las realidades objetivas de las campañas y los candidatos (Chihu, 2016; Panagopoulos & Wang, 2025; Tlapalal, 2024). El modelo del actor racional planteaba que los votantes recopilan y jerarquizan la información política recibida desde la comunicación política. Sin embargo, se ha demostrado que los actores sociales no suelen tener ese comportamiento racional, puesto que son organismos cibernéticos y, por tanto, tienen necesidad de conservar recursos, jerarquizando necesidades y problemas, y priorizando la supervivencia. De esta manera, las decisiones

---

<sup>50</sup> En la fotografía de la nota de Regeneración titulada “AMLO cerrará campaña en el Estadio Azteca” se puede observar al ex Presidente de la República escribiendo en un cartel en forma de corazón (Regeneración, 2018).

electorales se encuentran a la cola de esa jerarquización. Lo anterior se da debido a que los beneficios electorales no son directos y los logros colectivos, por tanto, la acción individual del voto tiene poca incidencia en la acción colectiva; es decir, el individuo tiene muy poco control en los resultados electorales y dedica poco tiempo y esfuerzo -tanto energético como cognitivo- en el planteamiento de su voto (Chihu, 2016).

En este texto, la emocionalización de la política se concibe como un proceso mediante el cual los discursos políticos movilizan afectos colectivos para orientar la acción política. En este sentido, las emociones no se consideran como estados psicológicos individuales, sino prácticas sociales y discursivas que intervienen en la configuración de la cultura política (Ahmed, 2015). Esta perspectiva permite el análisis de la movilización política contemporánea, principalmente, en el contexto de las campañas electorales, como un fenómeno en el que interactúan emociones, narrativas morales y procesos de polarización afectiva.

## **La subjetividad en el contexto de la cuarta transformación**

La política mexicana contemporánea ha tenido una reconfiguración muy particular desde la campaña electoral realizada para los comicios acontecidos el 1 de julio de 2018: ha predominado la apelación a los sentimientos y a las emociones, ya no sólo en el contexto de las campañas

electorales, sino también después de haberse formado el nuevo gobierno. A pesar de que, debido al contexto político global, este cambio se ha dado a nivel nacional, es decir, en cierta medida los distintos partidos políticos han adoptado las emociones como su pilar discursivo, dicha reestructuración fue absorbida, preeminentemente, por Andrés Manuel López Obrador y su movimiento político plasmado en el partido político Movimiento de Regeneración Nacional (Morena) que se ha conducido en un constante crecimiento a nivel nacional (Moreno, 2021) y que se ha podido corroborar en las últimas elecciones de 2024 a nivel federal y subnacional.

Andrés Manuel López Obrador (Tepetitán, 1953) posee una larga trayectoria política. El licenciado en Ciencias Políticas y Administración Pública por la Universidad Nacional Autónoma de México (UNAM) se afilió al Partido Revolucionario Institucional (PRI)<sup>51</sup> poco después de finalizar sus estudios en 1976 -aunque no se titularía hasta 1987-, tras lo cual comenzó su carrera en la administración pública, que estuvo estrechamente relacionada con la política social (Illades, 2020), como fueron la dirección del Instituto

---

<sup>51</sup> El Partido Revolucionario Institucional (PRI) es un partido político que fue fundado en 1929 como el Partido Nacional Revolucionario (PNR) por el sonorenses Plutarco Elías Calles quien, junto con Álvaro Obregón, representaba al Grupo Sonora perteneciente, en primera instancia, a la facción revolucionaria carrancista o constitucionalista. Este partido se convirtió, de acuerdo con Giovanni Sartori (2005), en partido hegemónico-pragmático y controló el sistema de partidos durante casi setenta años.

Indigenista de Tabasco o la de Promoción Social del Instituto Nacional del Consumidor (López Obrador, s. f.). Más tarde, formó parte de la escisión del ala izquierda del PRI, encabezada por Cuauhtémoc Cárdenas, que formó el Partido de la Revolución Democrática (PRD) y, entre 1996 y 1999, figuró como presidente de dicho partido (Illades, 2020). Por su parte, entre los años 2000 y 2005 fue jefe de gobierno de la Ciudad de México, cargo al cual postuló por el PRD. Su gobierno se caracterizó por impulsar programas sociales y construir infraestructura, entre otras cuestiones (BBC News Mundo, 2018). Una vez detentado ese cargo político, se afianzó su figura política y realizó el salto a la contienda presidencial.

Antes de ganar los comicios del año 2018, López Obrador participó en las elecciones presidenciales de 2006 y 2012. En ambas, fue el candidato de la coalición de los partidos políticos de izquierda; en 2006 conformada por el PRD, el Partido del Trabajo (PT) y Convergencia, que se denominó “Coalición por el Bien de Todos” y en 2012, la alianza política “Movimiento Progresista” también formada por el PRD, el Partido del Trabajo (PT) y Movimiento Ciudadano (antes Convergencia). En la contienda de 2006 estuvo a muy poco de ganar: obtuvo el 35.31% de los votos, por detrás del 35.89% del panista<sup>52</sup> Felipe Calderón Hinojosa

---

<sup>52</sup> Perteneciente al Partido Acción Nacional (PAN), fundado en 1939 por Manuel Gómez Morín, en oposición a las políticas socialistas del entonces

(INE, s. f.-a); mientras que en el 2012 la diferencia fue más notoria: logró el 31.64% de los votos, cuando el priista Enrique Peña Nieto obtuvo el 38.15% (INE, s. f.-b). De la misma manera, las campañas electorales realizadas también difirieron de manera significativa, tal y como lo veremos a continuación.

En la campaña de 2006, Andrés Manuel se centró, principalmente, en su persona y liderazgo debido a la popularidad que había logrado a lo largo de su jefatura en el gobierno de la Ciudad de México, principalmente, a raíz del desafuero<sup>53</sup>, con el cual logró movilizar a una amplia ciudadanía en su defensa (Pérez, 2012). Sin embargo, también hizo alusión a una unión de la derecha: el PRI y el Partido

---

presidente Lázaro Cárdenas, principalmente la nacionalización del petróleo, llevado a cabo el 18 de marzo de 1938, así como la educación laica y gratuita y la reforma agraria, con una naturaleza católica y a favor del libre mercado.

<sup>53</sup> El desafuero de AMLO se dio en el año 2004, cuando la administración de la Ciudad de México y, en consecuencia, Andrés Manuel, fue imputado de haber quebrantado una orden judicial que exhortaba anular la construcción de una calle en una parcela que había sido expropiada por gobiernos precedentes. A pesar de que se detuvo la obra, se alegó que dicha interrupción fue tardía y se le acusó directamente a AMLO de la situación. Sin embargo, como éste era jefe de gobierno, gozaba del fuero parlamentario que le otorgaba inmunidad a cualquier proceso administrativo o penal. Debido a ello, el entonces presidente Vicente Fox instó al Congreso de la Unión a llevar a cabo un juicio de desafuero, de manera que AMLO pudiera ser investigado por la justicia. Esta acción fue muy repudiada por la sociedad, que salió a las calles en apoyo a López Obrador, tras lo cual, Fox canceló la operación.



Acción Nacional (PAN), catalogando a ambos partidos como parte del problema del país y, por ende, lo opuesto a los intereses de las y los mexicanos. De esta manera, aunque de forma sutil, realizó una crítica al modelo económico neoliberal y planteó que la pobreza era resultado, no de una cuestión natural, sino de los malos gobiernos, una vez más, aludiendo a los gobiernos priistas y panista anteriores. Asimismo, propuso la necesidad de disminuir el gasto público derivado de los excesivos sueldos de los servidores públicos.

Otro aspecto que prevaleció en su discurso fue la necesidad de luchar contra la corrupción en aras de lograr un aumento del gasto público con la finalidad de implementar políticas sociales. Así:

López Obrador se asumió como representante de la izquierda, cuestionando severamente a sus dos contrincantes y sobre todo a quien no dejó de llamar el “títere de la derecha”, el abanderado de Acción Nacional, Felipe Calderón. Una izquierda moderada, centrista y pragmática. (Reveles, 2008, p. 102)

Sin embargo, la campaña electoral negativa ejercida, sobre todo, desde el ala panista, contra AMLO, aunado a la falta de una estrategia de defensa de dichos ataques de la oposición, se derivaron en una disminución paulatina de la popularidad del candidato (Vivero, 2008). Según el entonces Instituto Federal Electoral (IFE), actualmente Instituto

Nacional Electoral (INE), Andrés Manuel fue derrotado por Felipe Calderón. Sin embargo, AMLO no aceptó los resultados del IFE, acusó fraude electoral, exigió el recuento de cada uno de los votos casilla por casilla (Méndez, 2006) y realizó un mega plantón en la Ciudad de México acompañado de mega marchas, con lo cual su posicionamiento ideológico se posicionó más hacia la izquierda.

En 2012, la campaña electoral de López Obrador, si bien continuó en la misma línea que en 2006 en cuanto a la austeridad administrativa, corrupción y en subrayar la necesidad de políticas sociales, dio un vuelco hacia el plano emocional. Esto ocurrió con la publicación fechada el 6 de diciembre de 2011 en el periódico *La Jornada* del artículo “Fundamentos para una república amorosa”. En ese texto, AMLO, planteó, entre otras cuestiones, lo siguiente:

La decadencia que padecemos se ha producido, tanto por la falta de oportunidades de empleo, estudio y otros satisfactores básicos como por la pérdida de valores culturales, morales y espirituales. Por eso nuestra propuesta para lograr el renacimiento de México tiene el propósito de hacer realidad el progreso con justicia y, al mismo tiempo, auspiciar una manera de vivir, sustentada en el amor a la familia, al prójimo, a la naturaleza y a la patria. (López Obrador, 2011)

Además, agregó:

La descomposición social y los males que nos aquejan, no sólo deben contrarrestarse con desarrollo y bienestar y medidas coercitivas. Lo material es importante, pero no basta: hay que fortalecer los valores morales. [...] el propósito es contribuir en la formación de mujeres y hombres buenos y felices, con la premisa de que ser bueno es el único modo de ser dichoso. [...] En suma, estos fundamentos para una república amorosa deben convertirse en un código del bien. De ahí que hagamos el compromiso de convocar con este propósito a la elaboración de una constitución moral. (López Obrador, 2011)

Por consiguiente, a raíz de la publicación de este texto, Andrés Manuel comenzó a introducir aspectos emocionales - sobre todo relacionados con la simbología religiosa- en su discurso, apelando a que una de las causas de la insatisfacción generalizada que existía en la sociedad mexicana era la pérdida de valores. De esta manera, comenzó a sugerir que la fórmula para poder recuperar la felicidad de las y los ciudadanos mexicanos/as partía de la consolidación de los valores morales basados en “el apego a la verdad, la honestidad, la justicia, la austeridad, la ternura, el cariño, la no violencia, la libertad, la dignidad, la igualdad, la fraternidad y a la verdadera legalidad” (López Obrador, 2011). Es el discurso que manejará posteriormente para las campañas presidenciales de 2012 y, con mayor énfasis, de 2018, así como a lo largo de su sexenio (2018-2024).

Como se ha mencionado con anterioridad, en la campaña presidencial de 2012, AMLO fue el candidato de la coalición “Movimiento Progresista”. Sin embargo, cabe destacar que, para el 2011, ya se había fundado el Movimiento de Regeneración Nacional (Morena) que se había conformado como Asociación Civil y que en ese contexto tuvo la finalidad de acaparar el movimiento político y social que se había generado en torno a la figura de López Obrador y dirigirlo al logro de unos buenos resultados electorales del candidato en 2012 (Martínez, 2014). Por tanto, si bien, Andrés Manuel no logró el porcentaje de votos esperado, esa campaña le sirvió para ir consolidando y ampliando las bases sociales que potencialmente podrían votar por él y Morena, una vez constituido como partido político en 2014, en futuras elecciones. Así, el principal *slogan* de su campaña realizada en 2012 fue el “llamado a realizar un ‘cambio verdadero’ entre todos los mexicanos para consolidar una nueva República justa, humana, digna, libre, democrática y amorosa” (Chihu, 2016, p. 121). En su primer acto de campaña realizado en su municipio natal, aseveró que se comprometía a entregar su corazón al pueblo de México (Gil & Vergara, 2012). Por ende, en esa campaña presidencial prevalecieron las alusiones a los sentimientos, junto con la aseveración de que él personificaba la única alternativa a la corrupción, deshonestidad e injusticia que impregnaba al país.

La campaña de 2018 fue muy similar a la del 2012, sin embargo, la disimilitud fundamental radica en la existencia

de un nuevo partido, Morena, que se había construido en torno a la figura política del propio Andrés Manuel y que desde hacía siete años había estado trabajando en el fortalecimiento del respaldo de las bases sociales a ese movimiento político que defiende “el amor al prójimo y a la patria” (Martínez, 2014, p. 210) y cuyo eslogan evoca directamente a una de las emociones más importantes del ser humano: la esperanza<sup>54</sup>. De la misma forma, otra diferencia es la coalición realizada, no sólo con el partido de izquierda PT, sino con el Partido Encuentro Social (PES), un partido político caracterizado por propugnar a favor de las familias tradicionales y por posicionarse en contra el aborto, de una clara influencia cristiana y de ideología ubicada en el centroderecha (Siscar, 2014). Esos tres partidos formaron la coalición “Juntos Haremos Historia”, de la cual, Andrés Manuel, figuró como candidato (INE, 2018).

Otra de las características de la campaña y precampaña electoral realizada por López Obrador para las elecciones de 2018 fue que recorrió cada rincón del país escuchando las necesidades, inquietudes, opiniones y sensaciones de la ciudadanía mexicana, mostrando una gran empatía hacia ella, lo cual también influyó en el auge de su popularidad, mientras que, al mismo tiempo, permitió al entonces mandatario conocer de primera mano la situación del país y poder generar estrategias para mejorar, primero, su campaña y,

---

<sup>54</sup> El lema de Morena es “La esperanza de México”.

posteriormente, su gobernanza, pero, por supuesto, su discurso político. Lo anterior lo hizo, además, pregonado que no había realizado gastos excesivos en esos viajes y transmitiendo la imagen de un político cercano al ciudadano, el cual llega en una camioneta en lugar de un avión privado o en un auto austero en vez de un lujoso vehículo. El éxito de esa acción fue tal que, posteriormente, figuras como Ricardo Anaya replicaron esta estrategia durante los años siguientes, realizando recorridos similares por el país a través de distintos municipios con el objetivo de conocer de primera mano los problemas que padecían los ciudadanos mexicanos; un ejemplo de ello fue el recorrido que realizó junto con una enfermera mexiquense en el camino que ella debía realizar diariamente para llegar hasta su trabajo (Jiménez, 2020).

En cuanto al discurso político de las elecciones presidenciales de 2018, AMLO se centró en la anticorrupción y la austeridad (Ahuactzin, 2020), pero también en el rechazo al Pacto por México<sup>55</sup>, desmarcándose, así, de los demás

---

<sup>55</sup> El Pacto por México se puede considerar como una de las alianzas más importantes que se ha dado en los últimos años en el marco de la política mexicana. Este acuerdo permitió la alianza política entre el PAN y PRD con el gobierno de Enrique Peña Nieto, del PRI. De esta manera, el presidente pudo implementar las políticas que priorizaba, mientras que al mismo tiempo fortalecía el poder presidencial (Espinoza, 2015). Con este pacto se generó una agenda común entre distintos partidos superando, así, los desacuerdos que habían existido en los gobiernos previos y la aprobación de las iniciativas llevadas a cabo por el presidente Enrique Peña Nieto por un Congreso en el que el PRI no tenía la mayoría suficiente para respaldar a su candidato.

partidos políticos mexicanos, incluyendo al mismo PRD en el que había militado desde sus inicios -de hecho, uno de los principales pilares de la campaña fue la equiparación de los partidos PRI y PAN como un mismo proyecto político, catalogándolo como “PRIAN”-. A su vez, a lo largo de su campaña subrayó que era la única alternativa de un cambio real en el país y que con su triunfo se daría inicio a una nueva era en la política mexicana, a la cual hacía alusión el mismo nombre de la coalición. Sin embargo, la piedra angular del discurso político de la campaña 2018 de AMLO fueron los sentimientos. Comenzando con la etiqueta #AMLOVE utilizada en las redes sociales y que para mayo de ese año se convirtió en tendencia en Twitter, la cual, incluso más tarde pasó a ser una marca registrada (Bárceñas, 2018), y continuando con frases que mencionó como “Abrazos, no balazos”, que inclusive se convertiría en un término para definir la política pública en materia de seguridad del gobierno de Andrés Manuel y del proyecto político de la Cuarta Transformación.

En este sentido, buena parte de su discurso tiene similitudes con los postulados de la Teología de la Liberación latinoamericana que, de forma abreviada se puede definir como:

una reflexión que, a partir de la praxis y dentro del ingente esfuerzo de los pobres, junto con sus aliados, busca en la fe cristiana y en el Evangelio de Jesucristo

la inspiración para el compromiso contra la pobreza y en pro de la liberación integral de todo hombre. (Concha, 1997, p. 1558)

Así lo demuestran la bandera de “primero los pobres”, el grito de “viva el amor al prójimo” enarbolado, ya como mandatario, en el Grito de la Independencia del 2020 - aunque al día siguiente haya negado que se relacionara con el cristianismo como tal («¿Por qué AMLO gritó “viva el amor al prójimo” en el Grito de Independencia? Él lo explica», 2020)-, o el perdón otorgado a los ex presidentes y el sugerido para los criminales de la delincuencia organizada. De esta manera, AMLO pregonó que los altos índices de violencia que se padecieron a lo largo de la historia reciente del país se podían mitigar o erradicar a través de un discurso moralizante con el que la criminalidad desaparecería en consecuencia o que todos los ciudadanos se volverían honestos gracias al ejemplo del presidente. De hecho, el sexenio de Andrés Manuel López Obrador fue el más violento de la historia de México con 202 mil asesinatos en el país (Espino, 2025).

Por ende, la situación del país mejoraría gracias a una narrativa moral y no a la implementación de determinadas políticas encaminadas a resolver los problemas estructurales que padece México en la era contemporánea, como la gran desigualdad económica, la corrupción en los diversos niveles del estado y el rezago educativo, entre otros. En suma, la mejora de la situación nacional se planteó no sólo en términos



de políticas públicas concretas, sino también y, sobre todo, mediante una narrativa moral que atribuía la transformación social al fortalecimiento de valores éticos y comunitarios. Andrés Manuel transmitió que el escenario mexicano mejoraría gracias a una Cartilla Moral que todos los ciudadanos adoptarían y, de esta manera, por mera convicción, se lograría la transformación del país. Dicha Cartilla plantea que “la conducta moral, esto es, movida por el bien, nos permite vivir en paz con nosotros mismos y en armonía con los demás” (Martínez, 1992, p. 8) y, por su parte, “La práctica del bien, objeto de la moral, supone el acatamiento a una serie de respetos [...] Estos respetos equivalen a los “mandamientos” de la religión” (Martínez, 1992, p. 11). Ese discurso tuvo enorme calado en buena parte de la ciudadanía que, harta y decepcionada por los resultados obtenidos tras los últimos sexenios panistas y priista, optó por una vía distinta y alejada de los estándares políticos tradicionales.

Esto, en buena medida, se puede explicar a través de la influencia que tienen las emociones y los afectos en las personas y que generan acciones determinadas, como puede ser la determinación de votar por una figura política diferente que transmite confianza y genera esperanza, así como las características particulares del país. Pues, a pesar de que va en aumento la población de 5 años y más que considera no pertenecer a ninguna religión, el último censo realizado en 2020 ha demostrado que la población adscrita a la religión

católica sigue siendo muy alta (el 77.7%) y que la población que profesa el protestantismo o el cristianismo-evangélico va al alza (11.2%), por lo que se ha posicionado como la única religión que ha logrado un aumento de creyentes (Olvera & Martínez, 2021). Esto ayuda a entender por qué el discurso de Andrés Manuel que combinó ambas creencias, tanto la cristiana como la católica -puesto que también ha pregonado su veneración por la Virgen de Guadalupe-, tuvo tanta influencia en la sociedad mexicana. Pero, en general, la alusión que el mandatario realizó de forma recurrente a la religiosidad y a lo simbólico como fuerza de cambio político o de ayuda ante diversas adversidades y que generó un importante impacto emocional en la ciudadanía, constituyendo la principal fuerza de la movilización política del proyecto político de Morena.

## **Conclusión**

A lo largo de este texto se ha demostrado que el sexenio de Andrés Manuel López Obrador (2018–2024) constituye un caso ejemplar para comprender el papel que tienen las emociones como mecanismo de movilización política en la política contemporánea. A través de un marco teórico que reconoce la centralidad de la subjetividad, se observó que los ciudadanos no actúan únicamente a partir de cálculos racionales, sino de percepciones, afectos, significados simbólicos y disposiciones morales construidas social y discursivamente. En este sentido, la movilización emocional

durante la campaña electoral de 2018 y el sexenio de AMLO operó como un elemento fundamental para articular apoyos, configurar identidades políticas y producir formas de pertenencia y diferenciación. El análisis de las campañas de 2006, 2012 y 2018 mostró que la estrategia discursiva de las campañas de AMLO transitó de una retórica centrada en la injusticia y el antagonismo político hacia una narrativa afectiva basada en valores morales, símbolos religiosos, apelaciones al amor y la esperanza, y una reformulación ética de la vida pública. Este giro se consolidó durante el sexenio de Andrés Manuel, donde la comunicación presidencial - especialmente a través de las conferencias matutinas conocidas como “mañaneras”- reforzó la dimensión emocional que dividió a la ciudadanía entre “pueblo bueno” y adversarios morales, contribuyendo a una polarización afectiva sostenida.

Asimismo, se evidenció que esta emocionalización no puede entenderse sin considerar el contexto sociopolítico previo: el desencanto con los partidos tradicionales, la percepción de corrupción estructural, los altos niveles de violencia y la creciente demanda de soluciones simbólicas más que programáticas. Todo ello configuró un terreno fértil para que la Cuarta Transformación fuera interpretada como una promesa de transformación y ruptura política en el terreno afectivo, más que como un proyecto de reforma institucional tradicional. De esta manera, la estrategia política de López Obrador fue políticamente eficaz: permitió

movilizar a amplios sectores sociales, consolidar un movimiento de largo alcance y producir una narrativa afectiva resistente a la crítica tecnocrática o programática. Sin embargo, también generó tensiones importantes: profundizó la polarización, desplazó debates racionales sobre políticas públicas y favoreció la idea de que la transformación nacional podía basarse en cambios morales antes que estructurales.

Así, a partir del análisis realizado, pueden identificarse tres hallazgos principales. El primero corresponde a la movilización política asociada al proyecto de la Cuarta Transformación que se articuló, en buena medida, a través de narrativas afectivas que construyeron identidades políticas basadas en valores morales y símbolos compartidos. El segundo, dichas narrativas contribuyeron a la consolidación de una polarización afectiva que reconfiguró el debate público en términos morales en lugar de programáticos. El tercero, el caso mexicano muestra que las emociones pueden operar como recursos políticos estratégicos en contextos de desafección democrática y crisis de representación. Estos elementos sugieren la necesidad de desarrollar investigaciones empíricas que analicen de manera sistemática los mecanismos mediante los cuales las emociones intervienen en los procesos de movilización política en las democracias contemporáneas.

A partir de lo anterior, se abren varias líneas de investigación relevantes. Entre ellas destacan: ¿cómo operan

las emociones en contextos de posdemocracia y saturación informativa? ¿Qué formas de subjetividad política se consolidaron durante el sexenio de la 4T y cómo perduran en el escenario posterior a 2024? Finalmente, ¿hasta qué punto la emocionalización de la política se ha vuelto una característica estructural de las democracias latinoamericanas contemporáneas? Responder estas preguntas permitirá profundizar en la comprensión de un fenómeno que no se limita al caso mexicano, sino que constituye una tendencia global: la creciente centralidad de las emociones en la disputa por el poder político y en la configuración de la vida pública.

## Referencias

- Ahmed, S. (2015). *La política cultural de las emociones*. UNAM.
- Ahuactzin, C. E. (2020). Política y persuasión en México. Construcción del ethos en los spots electorales de Andrés Manuel López Obrador en 2018. *Discurso & Sociedad*, 14(1), 59–85.  
<https://doi.org/10.14198/dissoc.14.1.4>
- Ahuactzin, C. E., & González, R. A. (2019). *Elecciones México 2018. Entre la comunicación y la polarización*. Benemérita Universidad Autónoma de Puebla.  
[https://biblioteca.clacso.edu.ar/Mexico/icgde-buap/20210219022038/Elecciones\\_Mexico\\_2018.pdf](https://biblioteca.clacso.edu.ar/Mexico/icgde-buap/20210219022038/Elecciones_Mexico_2018.pdf)

- Aira, T. (2020). *La política de las emociones. Cómo los sentimientos gobiernan el mundo*. Arpa & Alfil Editores.
- Amador, J. (2014). *Comunicación y cultura. Conceptos básicos para una teoría antropológica de la comunicación*. UNAM.
- Bárcenas, X. (2018, 6 de noviembre). AMLOVE ya es marca registrada y ellos son los «dueños». *El Sol de México*.  
<https://www.elsoldemexico.com.mx/mexico/sociedad/amlove-ya-es-marca-registrada-y-ellos-son-los-duenos-2621060.html>
- BBC News Mundo. (2018, 3 de julio). *Qué hizo bien (y qué mal) López Obrador cuando gobernó la Ciudad de México*. Animal Político.  
<https://www.animalpolitico.com/2018/07/cuando-amlo-gobierno-cdmx/>
- Bisquerra, R. (2017). *Política y emoción*. Pirámide.
- Cantero, C. (2018). *Sociedad digital. Razón y emoción*. Editorial UOC.
- Chihu, A. (2016). *La imagen de los candidatos presidenciales (México 2012)*. Universidad Autónoma Metropolitana-Miguel Ángel Porrúa.
- Concha, M. (1997). Teología de la liberación. En N. Bobbio, N. Matteucci, & G. Pasquino (Eds.), *Diccionario de política* (pp. 1557–1563). Siglo XXI.
- Crouch, C. (2004). *Posdemocracia*. Taurus.

- Crouch, C. (2020). *Post-democracy after the crises*. Polity Press.
- Echeverría, B. (2011). *Ensayos políticos*. Ministerio de Coordinación de la Política y Gobiernos Autónomos Descentralizados.
- Espino, M. (2025, 1 de agosto). Sexenio de AMLO, el más violento de la historia: Inegi; registra 202 mil asesinatos en el país. *El Universal*. <https://www.eluniversal.com.mx/nacion/sexenio-de-amlo-el-mas-violento-de-la-historia-inegi-registra-202-mil-asesinatos-en-el-pais/>
- Espinoza, R. (2015). El fortalecimiento del Presidente de la República en el Pacto por México. En H. Zamitis (Ed.), *Pacto por México, agenda legislativa y reformas 2013-2014*. UNAM-La Biblioteca.
- Ferrajoli, L. (2014). Crisis económica, colapso de la democracia. En L. Salazar (Ed.), *¿Democracia o posdemocracia? Problemas de la representación política en las democracias contemporáneas* (pp. 47–53). Ediciones Coyoacán.
- Frijda, N. (2012). Emotions and actions. En A. S. R. Manstead, N. Frijda, & A. Fischer (Eds.), *Feelings and emotions. The Amsterdam Symposium* (pp. 158–173). Cambridge University Press.
- Gil, J., & Vergara, R. (2012, 30 de marzo). Ofrece López Obrador entregar su corazón y serenar al país. *Proceso*. <https://www.proceso.com.mx/nacional/politica/2012>

[/3/30/ofrece-lopez-obrador-entregar-su-corazon-serenar-al-pais-100748.html](#)

Illades, C. (2020). *Vuelta a la izquierda: La cuarta transformación en México: Del despotismo oligárquico a la tiranía de la mayoría*. Océano.

INE. (s. f.-a). *Elección de Presidente de los Estados Unidos Mexicanos*. Recuperado 16 de marzo de 2026, de [https://portalanterior.ine.mx/documentos/Estadisticas2006/presidente/gra\\_nac.html](https://portalanterior.ine.mx/documentos/Estadisticas2006/presidente/gra_nac.html)

INE. (s. f.-b). *Las elecciones del primero de julio: Cifras, datos, resultados*. Recuperado 16 de marzo de 2026, de <https://portalanterior.ine.mx/archivos3/portal/historico/recursos/IFE-v2/CNCS/CNCS-IFE-Responde/2012/Julio/Le010712/Le010712.pdf>

INE. (2018, 16 de marzo). *Recibe INE solicitud de registro de Andrés Manuel López Obrador como candidato a la Presidencia de la República*. Central Electoral. <https://centralelectoral.ine.mx/2018/03/16/recibe-ine-solicitud-de-registro-de-andres-manuel-lopez-obrador-como-candidato-la-presidencia-de-la-republica/>

Jiménez, H. (2020, 3 de marzo). *Anaya se sube a una combi y al Metro para acompañar a enfermera*. *El Universal*. <https://www.eluniversal.com.mx/nacion/ricardo-anaya-se-sube-una-combi-y-al-metro-para-acompanar-enfermera>



- Lechner, N. (1988). *Los patios interiores de la democracia. Subjetividad y política*. FLACSO.
- López Obrador, A. M. (s. f.). *Lic. Andrés Manuel López Obrador*. Recuperado 16 de marzo de 2026, de <https://lopezobrador.org.mx/semblanza/>
- López Obrador, A. M. (2011, 6 de diciembre). Fundamentos para una república amorosa. *La Jornada*. <https://www.jornada.com.mx/2011/12/06/opinion/009a1pol>
- Maffesoli, M. (2013). *El ritmo de la vida. Variaciones sobre el imaginario posmoderno* (Digital). Siglo XXI.
- Magaña, I. (2016). *Contribuciones a la psicología política en América Latina. Contextos y escenarios actuales*. RIL editores.
- Martínez, J. L. (1992). *Cartilla moral*. Secretaría de Educación Pública. <https://lopezobrador.org.mx/wp-content/uploads/2019/01/CartillaMoral.pdf>
- Martínez, J. P. (2014). Historia del partido Movimiento de Regeneración Nacional (Morena). *Revista de Derecho Estasiológico «Ideología y militancia»*, (4), 195–217.
- Méndez, E. (2006, 1 de agosto). Quienes hicieron el fraude, culpables del plantón: AMLO. *La Jornada*. <https://www.jornada.com.mx/2006/08/01/index.php?section=politica&article=005n1pol>
- Moreno, A. (2021, 4 de febrero). Morena aventaja... con apoyo de adultos mayores, universitarios y personas con menos estudios. *El Financiero*.

<https://www.elfinanciero.com.mx/nacional/con-menos-estudios-y-mas-edad-el-mayor-apoyo-a-4t>

Moreno, A. (2024). Facetas de la polarización política en México. *Foro Internacional*, 64(2), 189–238.

<https://doi.org/10.24201/fi.v64i2.3064>

Mouffe, C. (1999). *El retorno de lo político. Comunidad, ciudadanía, pluralismo, democracia radical*. Paidós.

Olvera, G., & Martínez, A. (2021, 25 de enero). Catolicismo y otras religiones pierden creyentes en México. *Milenio*.

<https://www.milenio.com/politica/comunidad/catolicismo-pierde-creyentes-censo-inegi-2021>

Panagopoulos, C., & Wang, C.-H. (2025). Rationality, affect, and vote choice. *Frontiers in Political Science*, 1–13.

<https://doi.org/10.3389/fpos.2025.1640408>

Paramio, L. (2002). La globalización y el malestar de la democracia. *RIFP*, 20, 5–23.

Pérez, G. (2012). El cambio de preferencias electorales durante la campaña para la elección de Presidente de la República en el proceso electoral de 2006 en México. En *El cambio racional de preferencias en el proceso electoral de 2006 en México. Una aproximación a las teorías de la elección racional en la ciencia política* (pp. 138–191). Cámara de Diputados del H. Congreso de la Unión.

¿Por qué AMLO gritó "viva el amor al prójimo" en el Grito de Independencia? Él lo explica. (2020, 17 de septiembre). *El Universal*.

<https://www.eluniversal.com.mx/nacion/amlo-por-que-grito-viva-el-amor-al-projimo-en-el-grito-de-independencia-el-lo-explica>

Rancière, J. (1996). *El desacuerdo. Política y filosofía*. Ediciones Nueva Visión.

Regeneración. (2018, 6 de junio). *AMLO cerrará campaña en el Estadio Azteca*. <https://regeneracion.mx/amlo-cerrara-su-campana-en-el-estadio-azteca-habra/>

Reveles, F. (2008). Los proyectos políticos de los candidatos presidenciales en el 2006: Las propuestas de gobierno. En M. Larrosa & P. J. Becerra (Eds.), *Elecciones y partidos políticos en México, 2006* (pp. 77–112). Universidad Autónoma Metropolitana Unidad Iztapalapa.

Rocha, R. (2005). La subjetividad política de los mexicanos y la construcción del comportamiento democrático. En J. O. Manjarrez, H. Oudhof, & J. L. Gama (Eds.), *Psicología social y cultura política* (pp. 73–102). Universidad Autónoma del Estado de México.

Sartori, G. (2005). *Partidos y sistemas de partidos. Marco para un análisis*. Alianza.

Siscar, M. (2014, 15 de julio). *Partido Encuentro Social, a favor de las familias tradicionales y contra el aborto*. Animal Político.  
<https://www.animalpolitico.com/2014/07/partido-encuentro-social-favor-de-las-familias-tradicionales-y-contra-el-aborto/>

- Tlapalal, D. C. (2024). Las emociones en el discurso político y la decisión de voto. *Contextualizaciones Latinoamericanas*, 1(32), 149–167. <https://doi.org/10.32870/cl.vii32.8092>
- Vivero, I. (2008). El Partido de la Revolución Democrática en las elecciones presidenciales de 2006. Notas sobre la campaña electoral. En M. Larrosa & P. J. Becerra (Eds.), *Elecciones y partidos políticos en México, 2006* (pp. 77–112). Universidad Autónoma Metropolitana Unidad Iztapalapa.
- Wagner, W., Hayes, N., & Flores Palacios, F. (2011). *El discurso de lo cotidiano y el sentido común. La teoría de las representaciones sociales*. Anthropos.
- Žižek, S. (2010). *En defensa de la intolerancia*. Diario Público.

# Itinerarios del pensamiento político y las hegemonías en la configuración del sujeto social

*Itineraries of Political Thought and Hegemonies in the Configuration of the Social Subject*

**Miguel Zavala Reyes<sup>56</sup>**

## Resumen

Este capítulo examina los itinerarios del pensamiento político desde la polis griega hasta las formas comunales contemporáneas, mostrando cómo las continuidades y rupturas han configurado al sujeto social en relación con distintas hegemonías políticas. La polis, como matriz

---

<sup>56</sup> Instituto Venezolano de Investigaciones Científicas IVIC Caracas.  
<https://orcid.org/0000-0002-5540-119X>

originaria, abrió el debate sobre ciudadanía, justicia y poder, mientras el tránsito hacia el Estado moderno consolidó categorías como soberanía y representación. Las ideologías democráticas renovadas y las epistemologías del sur amplían hoy la discusión, proponiendo alternativas frente a las crisis de legitimidad y representación. En este marco, la posverdad, las *fake news* y la desconfianza en la ciencia se convierten en desafíos centrales para la democracia y la difusión del conocimiento. La propuesta de este trabajo es ofrecer una narrativa neutral, objetiva y pedagógica que articule ética, epistemología social y pensamiento político contemporáneo, con el propósito de proyectar un modelo plural y emancipador del sujeto social en el siglo XXI.

**Palabras clave:** Polis; hegemonía; Estado; sujeto social; epistemología; democracia; emancipación; posverdad.

## **Abstract**

This chapter examines the itineraries of political thought from the Greek polis to contemporary communal forms, highlighting how continuities and ruptures have shaped the social subject in relation to political hegemonies. The polis, as an origin matrix, opened the debate on citizenship, justice, and power, while the transition to the modern State consolidated categories such as sovereignty and representation. Renewed democratic ideologies and the epistemologies of the South expand today's discussion,

offering alternatives to the crisis of legitimacy and representation. Within this framework, post-truth, fake news, and distrust in science emerge as central challenges for democracy and the dissemination of knowledge. The proposal of this work is to provide a neutral, objective, and pedagogical narrative that articulates ethics, social epistemology, and contemporary political thought, aiming to project a plural and emancipatory model of the social subject in the twenty-first century.

**Keywords:** Polis; hegemony; State; social subject; epistemology; democracy; emancipation; post-truth.

## **Introducción**

El estudio de los itinerarios del pensamiento político no constituye una mera recopilación historiográfica; representa una exploración de las estructuras teóricas que han moldeado la identidad del sujeto social en Occidente. En este sentido se analizan dichas trayectorias bajo el lente de la hegemonía, entendiendo que las categorías como justicia, poder y ciudadanía no son conceptos estáticos, sino construcciones en permanente disputa que determinan la posición del individuo frente a lo colectivo. El propósito fundamental es rastrear cómo los distintos regímenes de verdad han configurado la agencia del sujeto, transitando desde la participación directa en el espacio público hasta las actuales crisis de representación y legitimidad.

En este contexto, surge la posverdad como un fenómeno crítico que desafía los cimientos de la deliberación democrática. Más que una simple proliferación de datos inexactos, la posverdad se aborda en esta investigación como una posible estrategia de fragmentación discursiva que compromete la cohesión del sujeto social. Al erosionar el suelo común de la racionalidad compartida, este fenómeno no solo altera la comunicación política, sino que corre el riesgo de desarticular la capacidad de acción coordinada, desplazando el diálogo hacia lógicas de antagonismo donde el consenso se vuelve esquivo. Esta tesis se plantea como el eje articulador que conecta las herencias del pensamiento clásico con las tensiones de la contemporaneidad.

Para dotar a esta reconstrucción de un rigor operativo, la narrativa se organiza en torno a nodos críticos de transformación hegemónica. El itinerario comienza en la polis griega, analizada como la matriz donde la palabra y la verdad pública (*parresía*) fundamentaron la ciudadanía. Posteriormente, se examina el tránsito hacia el Estado moderno, nodo en el cual la razón institucional y el método científico buscaron estabilizar un orden soberano. Este recorrido se extiende hacia las ideologías democráticas renovadas y, finalmente, hacia las epistemologías del sur, integrando estas últimas como una respuesta necesaria ante las exclusiones de la racionalidad moderna. Cada transición se justifica por la crisis del paradigma anterior, permitiendo



una comprensión sistémica de cómo el poder se reconfigura ante los cambios sociales.

Finalmente, se integran diversas dimensiones para analizar la crisis actual: desde la ética de la comunicación hasta la influencia de las tecnologías digitales en lo que algunos autores denominan *infocracia*. Al contrastar estas perspectivas, se busca ofrecer un análisis profundo sobre la configuración del sujeto social en el siglo XXI. El objetivo no es solo describir una evolución de ideas, sino identificar las herramientas teóricas que permitan recuperar una agencia política sólida frente a los mecanismos de aislamiento y fragmentación que caracterizan el escenario político global actual. Nos adentramos en una lectura reflexiva orientada a repensar horizontes que permitan visibilizar la construcción de modelos plurales, sustentados en la autodeterminación de los pueblos y en un Estado de derecho que se erija como referente ético y emancipador del sujeto social en el siglo XXI.

## **Fundamentación metodológica**

Lejos de ser una recopilación exhaustiva, esta investigación se estructura como una reconstrucción operativa de los itinerarios políticos bajo una metodología histórico-epistemológica. La coherencia del análisis y la unidad del constructo se aseguran mediante tres directrices de ensamblaje que rigen el estudio:

- **Criterios de selección y muestreo teórico.** La elección de autores y periodos no responde a una linealidad cronológica accidental, sino a su relevancia en la configuración de "regímenes de verdad". Se han seleccionado cuatro nodos críticos donde la hegemonía política se reconfigura para dar respuesta a crisis de identidad y orden: la **polis griega** (como matriz de la verdad pública), el **Estado moderno** (como institucionalización de la razón), las **ideologías democráticas renovadas** (frente a la crisis de la representación) y las **epistemologías del sur** (como emergencia de la justicia epistémica).
- **Lógica de tránsito agónico.** El tránsito entre estos bloques se justifica a partir de la insuficiencia de los paradigmas precedentes. La transición de la polis al Estado moderno se analiza como el desplazamiento de la virtud ética hacia la necesidad de una soberanía centralizada capaz de gestionar la complejidad técnica. Del mismo modo, el giro hacia las epistemologías del sur se fundamenta en el agotamiento de la racionalidad moderna ante las realidades de exclusión contemporáneas. Este enfoque permite comprender que la teoría política evoluciona como un esfuerzo humano por resolver tensiones entre convivencia y poder.

- **Integración operativa de la epistemología social.** Esta propuesta integra conceptos de la epistemología social —como el valor del testimonio y la categoría de injusticia epistémica— para evaluar la calidad del lazo social. Estos principios funcionan como indicadores para analizar la **posverdad**: si el sujeto pierde la confianza en el testimonio compartido y sufre una desautorización de su conocimiento, su capacidad de agencia política se anula. Así, la metodología permite analizar el impacto de la fragmentación hegemónica no solo en las instituciones, sino en la subjetividad misma del ciudadano.

## **La polis griega como matriz originaria**

El itinerario del pensamiento político occidental se inicia en la Polis griega, donde Platón y Aristóteles sentaron las bases normativas y empíricas de la teoría política. Platón, en *La República*, buscó reconciliar la teoría moral con la práctica política mediante el diseño de una comunidad regida por la armonía entre funciones y saberes, legitimando la autoridad del gobierno filosófico en nombre del bien común (Delannoy & Mack, 2008; Rosillo Olmedo, 2020). En su propuesta, la ciudadanía se medía por la capacidad de integrarse a esa arquitectura moral y obedecer a la razón como principio del mando. Aristóteles, por su parte, desplazó el eje hacia la praxis deliberativa: llamó al ser humano *zoon*

*politikon* porque la vida plena requería pertenencia a la polis y ejercicio de virtud, entendida no como ornamento, sino como criterio de legitimidad institucional. La justicia — distributiva y correctiva— articulaba mérito, ley y medida de lo común, y con ello definía la pertenencia ciudadana (Knoll, 2017; López Hana, 2014). Entre ambos aportes pudiéramos preguntarnos: ¿qué hace ciudadano al ciudadano? Platón respondió con conocimiento ordenado y educación del alma; Aristóteles, con hábito de virtud y experiencia constitucional. Dos vías distintas para una misma exigencia: que lo político sea praxis orientada al bien común (Quispe Arroyo, 2019).

Este diseño fundacional tensionó las categorías individuo y comunidad. Platón intuía del deseo privado y subordinaba la autonomía al bien común; Aristóteles reconocía pluralidad de regímenes y admitía el disenso regulado por ley y costumbre, aunque advertía la fragilidad de la democracia cuando se desbordaba hacia el interés sin medida (Delannoy & Mack, 2008; López Hana, 2014). La virtud cumplía un papel doble: habilitaba el juicio público y limitaba el poder, porque nadie podía gobernar bien sin moderación y justicia. Así, la praxis política se ataba a la ética: el ciudadano no era solo quien votaba, sino quien se formaba en hábitos que permitían deliberar sobre fines, elegir medios y responder por consecuencias (Knoll, 2017). La autoridad encontraba su legitimidad en la intersección entre saber, ley y virtud, mientras la tensión entre justicia y poder se volvía

constante. La polis se erigió, así, como matriz originaria de las categorías que aún sostienen la teoría política.

Bajo esta óptica, la Polis pasa a ser la matriz de la verdad pública. En la Polis la verdad era el vínculo que relacionaba al ciudadano con lo colectivo, la posverdad actúa hoy como su antítesis, buscando precisamente disolver ese lazo mediante la fragmentación del sentido compartido.

Sin embargo, esa matriz tuvo márgenes que no podemos soslayar. Mujeres, esclavos y extranjeros fueron excluidos de la ciudadanía, testimonio de una libertad construida sobre silencios constitutivos (Fuentes Santibáñez, 2012). Desde las epistemologías del Sur, esa herida se lee como memoria insurgente: la crítica no busca negar el legado clásico, sino interrogar su universalidad, revelar las jerarquías que lo sostienen y ampliar la deliberación a saberes, cuerpos y experiencias históricamente desautorizadas (de Sousa Santos, 2011; de Sousa Santos, 2016; Meneses & Bidaseca, 2018). La interpelación feminista añade otra capa, Nussbaum (2000) propone ciudadanía como capacidad efectiva de hacer y ser, con garantías materiales y simbólicas que rompan el pacto entre excelencia moral y exclusión social; Bidaseca (2011) denuncia el paternalismo colonial que disfrazaba emancipación y devuelve agencia a las sujetas históricamente tuteladas. Así, la polis se convierte en punto de partida para pensar las ideologías renovadas y las formas comunales que hoy buscan superar las exclusiones fundacionales.

El tránsito de la polis al Estado moderno muestra cómo las tensiones fundacionales se prolongan en nuevas configuraciones. La cronología política, como recuerda Yépez (1988), señala que las monarquías arcaicas fueron reemplazadas por oligarquías, y que el poder pasó a los jefes de las grandes familias que poseían tierra, riqueza y armas. Platón y Aristóteles sentaron las bases de la teoría del Estado, pero la modernidad transformó esas categorías en soberanía, representación y ciudadanía. La ciencia política, como disciplina independiente, se nutrió de préstamos de la sociología, la psicología y la historia, buscando credibilidad científica y métodos empíricos para comprender el gobierno y la política (Chitty, 2008; Pérez, 1998). En este itinerario, la neutralidad y la pluralidad emergen como dilemas persistentes: la política es persuasión, convencimiento y liderazgo, pero también es pluralidad de concepciones de mundo que disputan legitimidad (Chitty, 2008).

La continuidad entre la polis y la democracia contemporánea radica en la idea de que la vida que satisface los requerimientos mínimos a plenitud necesita instituciones comunes y deliberación pública. Lo que cambia es el perímetro de lo común y la materialidad de la participación: gobiernos abiertos, co-creación, rendición de cuentas y espacios de decisión no meramente consultivos (Gómez Díaz de León, 2018; Rodríguez Casallas et al., 2024). En América Latina, esa promesa avanza y retrocede como un horizonte en disputa. En experiencias disimiles la participación se ha

reducido a mera retórica. Sin embargo, cuando se articula a través de un diseño institucional sólido y mecanismos de control social, puede transformarse en un poder ciudadano efectivo. (Briceño & Maingon, 2015). La pregunta por la legitimidad reaparece bajo la perspectiva de diversas configuraciones teóricas: Held (2006) cartografía la pluralidad democrática; Young (2000) insiste en la inclusión como condición de justicia política; Mouffe (1999) recuerda que la democracia no supera el conflicto, sino que lo civiliza; y Rancière (2006) advierte que el odio a la democracia surge cuando la igualdad irrumpe y desordena la distribución de “quién puede hablar”. La polis, como matriz originaria, persiste en estas discusiones, no para idealizarla, sino para devolverle su potencia crítica y expandir la democracia como práctica viva de igualdad, deliberación y responsabilidad colectiva.

## **El tránsito hacia el Estado moderno**

La conformación del Estado moderno constituye un proceso histórico y teórico marcado por tensiones entre convivencia y dominación, libertad y orden. Las formulaciones clásicas de Hobbes, Locke y Rousseau ofrecen los cimientos de esta discusión, situando la política en el terreno del contrato social y la legitimidad del poder. Hobbes, desde la experiencia de la guerra civil inglesa, defendió la necesidad de un soberano absoluto para garantizar la paz y evitar la anomia social (Sasan, 2021; Zhang, 2024). Locke, en

contraste, vinculó el contrato a la protección de los derechos naturales, especialmente la propiedad, y a la posibilidad de disolver gobiernos que traicionen la confianza de los gobernados (Sasan, 2021). Rousseau radicalizó la noción al proponer la voluntad general como fundamento de la soberanía, abriendo paso a una concepción democrática que inspiró la Revolución francesa (Zhang, 2024).

El tránsito hacia el Estado moderno se justifica operacionalmente por el agotamiento de la virtud ética como única garantía de orden. La hegemonía se desplaza hacia una razón institucional y científica. Este cambio es fundamental para entender la crisis actual: la posverdad no solo ataca hechos aislados, sino que intenta dinamitar este cimiento de racionalidad técnica sobre el cual se edificó la soberanía y la seguridad del sujeto moderno.

Este debate fundacional se proyecta hacia la modernidad como un entramado de tensiones irreconciliables. La paradoja de la democracia constitucional, señalada por Habermas, muestra cómo la normatividad ideal convive con prácticas políticas que la erosionan (Jessop, 2015). En la misma línea, Streeck (2011) advierte que el capitalismo democrático está atravesado por crisis recurrentes derivadas de la tensión estructural entre mercados y votantes. La modernidad política, lejos de estabilizarse, se configura como un campo de conflicto



permanente, donde las promesas de libertad se cruzan con las exigencias de orden y hegemonía.

La expansión del mundo conocido, el desarrollo del comercio internacional y el ascenso del poder real en ciertos Estados dieron lugar a doctrinas como el absolutismo, definido por la afirmación de una soberanía monárquica sin límites. Este contexto fue el terreno sobre el que se desplegó una de las más importantes revoluciones en el campo de las ideas: la emergencia de la ciencia política como disciplina autónoma. Maquiavelo, en *El Príncipe*, liberó a los gobernantes de la sujeción a principios morales, justificando los medios empleados para la consecución de fines superiores y otorgando al Estado un carácter de rango superior (Martínez, 1988). En contraste, Erasmo de Rotterdam defendió la aplicación de preceptos evangélicos en la vida política, mientras Tomás Moro imaginó en *Utopía* una república ideal gobernada por la razón y la tolerancia religiosa (Scheck, 2021). Estas visiones contrapuestas revelan que el tránsito hacia el Estado moderno no fue lineal, sino un entramado de propuestas que oscilaron entre la afirmación del poder absoluto y la búsqueda de comunidades igualitarias.

La Reforma añadió otra capa de complejidad, al desafiar la hegemonía de la Iglesia católica y abrir paso a nuevas formas de autoridad política y religiosa. Lutero defendió la sumisión incondicional a la autoridad, mientras Calvino propuso un régimen clerical que subordinaba la vida

civil a la disciplina eclesiástica (Wolin, 1953). Frente a estas posturas, Bodin elaboró una teoría de la soberanía indivisible en la persona del rey, mientras Althusius defendió el principio de soberanía popular y el derecho de resistencia frente al tirano. El absolutismo, consolidado en figuras como Richelieu y Luis XIV, encarnó la concentración del poder en nombre del orden y la paz civil, pero también generó resistencias que alimentaron el pensamiento constitucional y democrático (Salmon, 1996).

La Ilustración, con Montesquieu, Voltaire, Rousseau y Kant, sistematizó conceptos que aún hoy sostienen la teoría política moderna: división de poderes, crítica a la tiranía, contrato social y el ejercicio de la paz como valor universal. Montesquieu destacó la necesidad de leyes adaptadas a las condiciones sociales y naturales; Voltaire denunció los abusos de las instituciones políticas y religiosas; Rousseau formuló la voluntad general como principio de soberanía; y Kant defendió una federación mundial de Estados republicanos. Estas propuestas, aunque diversas, coincidieron en afirmar que la política debía organizarse en torno a la justicia, la igualdad y la libertad, valores que se convirtieron en pilares del Estado moderno.

La teoría crítica contemporánea ha problematizado estas herencias. Fraser (2008) cuestiona la suficiencia del marco estatal-nacional para garantizar derechos en un mundo interdependiente, mientras Mouffe (2005) subraya el

carácter agonístico de lo político, recordando que la democracia no elimina el antagonismo, sino que lo canaliza en formas institucionalizadas. Cerny (2010) describe la emergencia de un “neopluralismo transnacional”, donde el poder se dispersa en múltiples escalas y actores, desafiando la centralidad del Estado. En este escenario, el Estado de bienestar enfrenta presiones de globalización, crisis fiscales y transformaciones sociales, lo que obliga a repensar sus fundamentos (Gough, 2015; Papachashvili et al., 2023). El marco de la inversión social se presenta como alternativa para articular bienestar y productividad en el siglo XXI (Hemerijck et al., 2023).

Finalmente, se reabren debates contemporáneos sobre el contrato social. La construcción de nuevos pactos sociales, como señala Jessop (2015) y autores como (Loewe et al., 2024), reconocen la pluralidad de actores y escalas, así como la necesidad de integrar justicia social, sostenibilidad ambiental y legitimidad democrática. El tránsito hacia el Estado moderno, lejos de ser un proceso concluido, se revela como una trayectoria abierta, marcada por la tensión entre hegemonías y resistencias, entre la promesa de la libertad y la persistencia de la dominación.

## **Ideologías democráticas renovadas**

La discusión sobre las ideologías democráticas renovadas se sitúa en el cruce de crisis estructurales y

búsquedas de alternativas que, desde la socialdemocracia hasta las epistemologías del sur, intentan redefinir la relación entre Estado, mercado y ciudadanía. La socialdemocracia, surgida en Europa a finales del siglo XIX, se presentó como una propuesta moderada frente al marxismo clásico. Su estrategia inicial consistió en organizar políticamente a la clase trabajadora dentro de la democracia liberal, aceptando el sistema de economía de mercado, pero promoviendo la intervención pública para corregir sus deficiencias. Con el Congreso de *Bad Godesberg* en 1959, la socialdemocracia reconoció la obsolescencia del marxismo ortodoxo y adoptó un pragmatismo que buscaba humanizar el capitalismo y reforzar el Estado de bienestar (Daviault & Vinay, 1959; Sassoon, 1996). Este modelo alcanzó su auge entre 1945 y 1973, pero la crisis del petróleo y el colapso del consenso de bienestar en los años setenta marcaron su declive. Ante el avance del neoliberalismo en los ochenta, la socialdemocracia respondió con la llamada “tercera vía”, concebida por Giddens (1999) como un marco de pensamiento que intentaba trascender tanto la vieja socialdemocracia como el neoliberalismo. Sin embargo, como advierte Streeck (2011), más que superar el neoliberalismo, esta estrategia lo adaptó a nuevas condiciones, generando una tensión permanente entre la promesa de cohesión social y la fragilidad de los sistemas de protección. En este sentido, la socialdemocracia renovada se convirtió en correctivo parcial, pero también en vehículo de adaptación, redefiniendo el vínculo entre

ciudadanía y mercado bajo el signo de la competitividad global y no de la justicia redistributiva.

La democracia cristiana, por su parte, articula la noción de persona frente al Estado desde principios de dignidad, subsidiariedad y solidaridad. Sus antecedentes modernos se remontan al Partido Popular Italiano fundado por Luigi Sturzo tras la Primera Guerra Mundial, inspirado en la Doctrina Social de la Iglesia (Caldera, 1996). La enseñanza social católica, como recuerda Byron (2009), se fundamenta en la dignidad humana como núcleo de los derechos sociales y políticos, mientras que Pazos-Vidal (2019) subraya la ambivalencia de la subsidiariedad: puede ser límite al poder estatal, pero también mecanismo de fortalecimiento comunitario. La solidaridad, formulada en la encíclica *Sollicitudo Rei Socialis* de Juan Pablo II (1987), se convierte en principio ordenador de políticas públicas, invitando a asumir responsabilidades compartidas en un mundo marcado por la interdependencia. La propuesta democratacristiana de una “economía social de mercado”, impulsada por Müller Armack y aplicada por Erhard en la Alemania de posguerra, buscó combinar la eficiencia del mercado con la justicia social, equilibrando libertad económica y equidad (Gómez Peralta, 2012). Este modelo, que inspiró proyectos políticos en Europa y América Latina, muestra cómo la democracia cristiana intenta evitar tanto el estatismo absorbente como el individualismo atomizante, proponiendo un capitalismo con rostro humano.

El socialismo del siglo XXI emerge como respuesta ante el desgaste de aquellas teorías que en la praxis gubernamental no solucionaban las problemáticas reales y sentidas que la población demandaba. Con un discurso revolucionario se visualiza la democracia directa y participativa trascendiendo el formalismo burocrático representativo. Heinz Dieterich formuló esta propuesta en 1996, difundida posteriormente por Hugo Chávez en el Foro Social Mundial de 2005 (Chávez, 2005). Su modelo se sustenta en tres ejes: desarrollismo democrático regional, economía de las equivalencias y democracia participativa. A diferencia de los intentos socialistas del siglo XX, Dieterich (2007) no plantea un modelo único, sino una macroeconomía metodológica que articula bloques regionales de poder estatal y popular. La economía de valores, basada en el tiempo de trabajo como medida del valor, busca superar las asimetrías del mercado capitalista. En Venezuela, los consejos comunales se convirtieron en espacios de disputa y experimentación democrática, ampliando la participación, aunque tensionados por estructuras estatales centralizadas (Wilde, 2015). En Bolivia y Ecuador, la redistribución territorial y las autonomías reflejaron la búsqueda de hegemonías políticas legitimadas en la descentralización, aunque no siempre resolvieron las tensiones entre pluralidad social y control estatal (Stoessel, 2014). Como señala Lander (2014), estos procesos intentaron construir alternativas al neoliberalismo mediante justicia territorial y autogestión

comunal, aunque enfrentaron contradicciones entre autonomía comunitaria y articulación estatal.

Las epistemologías del sur complementan estas ideologías renovadas al cuestionar la universalidad de los modelos democráticos occidentales y reivindicar al sujeto social desde las luchas territoriales. De Sousa Santos (2014) denuncia el epistemicidio que supone imponer un único modelo de racionalidad política, mientras Escobar (2016) propone un “pensar-sentir con la Tierra” que vincula democracia y defensa de territorios. Grosfoguel (2011) muestra cómo la colonialidad persiste en las formas de representación política, invisibilizando sujetos y saberes. Estas perspectivas reconfiguran la noción de ciudadanía, que deja de ser un estatus jurídico abstracto para convertirse en práctica situada, vinculada a la defensa de territorios, cuerpos y comunidades.

## **Género, comunidad y emancipación del sujeto social**

La configuración del sujeto social en el siglo XXI exige reconocer las huellas de exclusión histórica que han marcado a mujeres y colectivos subalternos. Como advierte Quijano (2014), retomado por Mignolo (2007) y Walsh (2013), la colonialidad del poder no se limita a la economía y la política, sino que también se expresa en la epistemología que invisibiliza saberes y prácticas situadas. En este sentido, el feminismo comunitario de Julieta Paredes se erige como una

respuesta insurgente frente a la matriz patriarcal y neoliberal, al plantear que la emancipación no puede reducirse a la inclusión formal en estructuras que reproducen desigualdad, sino que debe articularse desde la comunidad como cuerpo político transformador (Paredes, 2015).

La ciudadanía, concebida como acción transformadora, se encuentra inmersa en un dinamismo que la redefine constantemente en función de su entorno social. Landau (2004) la entiende como un proceso histórico de construcción y reconstrucción, mientras Marshall (2005) la temporaliza en tres momentos: derechos civiles, derechos políticos y derechos sociales. Esta progresión revela que la ciudadanía no es un estatus fijo, sino una práctica que legitima la inclusión y la capacidad de los grupos sociales para disputar el sentido de la democracia. Sin embargo, frente al fenómeno de la exclusión, se hace necesario distinguir entre ciudadanía formal y sustancial, pues muchas personas poseen el reconocimiento jurídico sin gozar efectivamente de los derechos que este implica. La globalización, además, ha desplazado la ciudadanía más allá del Estado-nación, abriendo paso a categorías como la ciudadanía transnacional y supranacional, que cuestiona los límites tradicionales de pertenencia y participación.

La ciudadanía plural, lejos de ser un mero reconocimiento jurídico, implica cuestionar las bases mismas del modelo patriarcal que ha naturalizado la división entre lo



público y lo privado. Butler (2004) y Mohanty (2003) han mostrado cómo las narrativas universales del feminismo occidental tienden a homogeneizar experiencias, invisibilizando las luchas de mujeres racializadas e indígenas. Frente a ello, Calla (2010) subraya la emergencia de feminismos indígenas en América Latina, que reivindican la complementariedad y la memoria ancestral como fundamentos de una ciudadanía distinta, capaz de integrar la diferencia sin subsumirla en un universal abstracto.

La participación, entendida como acto individual y colectivo, se convierte en el núcleo de esta ciudadanía transformadora. Participar es integrarse a la vida de la comunidad de forma responsable, como derecho y deber, pero también como práctica que legitima el sistema democrático. Las ventajas de la participación —pluralidad, inclusión, responsabilidad social y generación de consensos— se enfrentan a resistencias como la exclusión, la defensa de intereses particulares y la cultura del individualismo acrítico. Superar estas resistencias exige un compromiso consciente y responsable, capaz de transformar la cultura política y abrir espacios de deliberación sustantiva. En este sentido, la participación social, a través de organizaciones, sindicatos y movimientos comunitarios, se convierte en instrumento de vertebración democrática y de disputa por un nuevo modelo de desarrollo inclusivo.

La comunidad y el territorio, como categorías insurgentes, se constituyen en espacios de resistencia frente a la lógica del capital y la colonialidad. Escobar (2016) reconoce que las luchas territoriales son también luchas ontológicas, pues defienden modos de existencia relacionales que el modelo extractivista busca anular. Walsh (2013) complementa esta perspectiva al señalar que los “otros saberes” no son meras alternativas locales, sino epistemologías críticas que cuestionan la hegemonía eurocéntrica y abren la posibilidad de un diálogo interepistémico. De Sousa Santos (2018) insiste en la necesidad de superar el “imperio cognitivo” para avanzar hacia una pluriversalidad que reconozca la coexistencia de múltiples mundos.

La comuna, entendida como horizonte emancipador, condensa la praxis viva de autogestión y justicia relacional. Federici (2012) recuerda que las luchas sociales han estado históricamente protagonizadas por mujeres, quienes han defendido la reproducción de la vida frente a los cercamientos del capital. En América Latina, esta defensa se traduce en experiencias de organización comunitaria que, como señala Grosfoguel (2011), buscan descolonizar tanto la economía como el pensamiento político, articulando prácticas de solidaridad y reciprocidad que desafían la lógica individualista. La comuna, entonces, no es solo una forma organizativa, sino una apuesta ética y política por reconstruir vínculos sociales desde la cooperación y la relacionalidad.

## **Epistemología social y ética en el itinerario político y la difusión del conocimiento**

El estudio de las condiciones colectivas que determinan la producción, circulación y legitimación del conocimiento —propio de la epistemología social— adquiere una dimensión política cuando se articula con la ética y con las disputas por la verdad en la esfera pública. En este cruce, la epistemología se transforma en itinerario político y filosófico, revelando cómo el saber no solo organiza la vida social, sino que también configura los marcos de legitimidad y emancipación en la praxis contemporánea. No se trata únicamente de analizar cómo se construye el saber, sino de reconocer que cada forma de enunciación, cada práctica de difusión y cada silenciamiento implican relaciones de poder que afectan directamente la justicia social y epistémica. En este sentido, el conocimiento deja de ser un recurso neutral para convertirse en un campo de batalla donde se negocian credibilidad, autoridad y legitimidad.

El testimonio ocupa un lugar central en este itinerario. Como forma de transmisión de experiencias y saberes, el testimonio puede ser fuente de verdad y resistencia, pero también puede derivar en patologías epistémicas como rumores, chismes y secretos, que producen lo verosímil más que lo verdadero (Adkins, 2017; Bertolotti, 2014; Ceriani Cernadas, 2017). Estas formas subterráneas de circulación del conocimiento configuran un “saber informal” que, aunque

marginal, influye en la construcción de imaginarios colectivos y en la percepción pública de los hechos. En la era digital, estas dinámicas se amplifican con las *fake news*, que operan como dispositivos de desinformación y posverdad, erosionando la confianza en las instituciones y generando un terreno fértil para la manipulación política y cultural (Quiroz Pacheco *et al.*, 2023). La posverdad, más que una simple mentira, es un régimen discursivo donde la emoción y la repetición desplazan a la evidencia, y donde la viralidad sustituye a la veracidad.

En este escenario, la injusticia epistémica emerge como una categoría crítica para comprender cómo ciertos grupos son sistemáticamente deslegitimados como portadores de conocimiento. La exclusión de voces, la invisibilización de experiencias y la producción deliberada de ignorancia informada constituyen mecanismos que perpetúan desigualdades y consolidan jerarquías epistémicas (Cohen & Garasic, 2024; Garzón Rodríguez & Acosta Navas, 2022). La injusticia epistémica no solo se manifiesta en la negación de credibilidad a los sujetos subalternos, sino también en la imposición de marcos interpretativos que limitan su capacidad de agencia y participación en la construcción de lo común.

Frente a estas tensiones, la ética de la participación se convierte en un principio rector. La investigación participativa y la ciencia ciudadana, al situar la

responsabilidad mutua como eje de colaboración, ofrecen un horizonte alternativo donde el conocimiento se construye desde la pluralidad y el cuidado (Groot & Abma, 2022; Groot et al., 2019). Este enfoque ético reconoce que la producción de saber no puede desligarse de las relaciones humanas que la sostienen, y que la justicia epistémica requiere prácticas de escucha, reconocimiento y corresponsabilidad. En paralelo, la organización en las comunidades se presenta como un modelo institucional que busca integrar el bien colectivo en las estructuras de acción compartida, resistiendo la privatización del conocimiento y promoviendo su circulación como recurso comunitario (Albareda & Sison, 2020).

En América Latina, la esfera pública digital se ha convertido en un espacio de disputa donde la democratización del acceso convive con la concentración de poder mediático y tecnológico. La expansión de Internet ha abierto posibilidades inéditas para la participación ciudadana, pero también ha revelado las tensiones entre inclusión y exclusión, entre apertura y control (Saavedra Utman & Santos, 2024). La política de la verdad en la región exige, por tanto, una mirada crítica que reconozca tanto las potencialidades emancipadoras de las redes como sus riesgos de captura por intereses corporativos y estatales.

La voz subalterna, en este itinerario, adquiere un protagonismo ineludible. No se trata de hablar por los otros, sino de abrir espacios para que los otros hablen desde sus

propias sensibilidades y experiencias. La descolonización de la investigación interpretativa, al atender a las sensibilidades subalternas y a la política de la voz, se convierte en un acto de resistencia frente a la hegemonía epistémica (Darder, 2017). En el ámbito del trabajo social, las voces disidentes de los usuarios de servicios cuestionan las estructuras que pretenden representarlos sin escucharlos, recordando que la justicia epistémica implica otorgar espacio y legitimidad a quienes históricamente han sido silenciados (Aaslund, 2024).

En este escenario, tan ampliamente descrito, siguiendo un recorrido temporal, se observa que, desde Aristóteles hasta Bodin, la política ha sido concebida como un orden racional y normativo; con la modernidad, se transformó en instituciones que regulaban la vida social. Hoy, en la era digital, el ejercicio político se redefine en el campo de la información, donde las *fake news* representan la crisis contemporánea de la verdad como fundamento del poder. Así, la evolución histórica de la política revela que su ejercicio siempre ha estado ligado a la construcción de narrativas, aunque ahora estas narrativas se disputen en un terreno marcado por la posverdad y la manipulación informativa.

Mouffe (2005) recuerda que lo político es siempre agonístico, y que la posverdad exacerba los antagonismos al desdibujar los límites entre lo verdadero y lo verosímil. Fraser (2008) añade que la justicia en un mundo globalizado requiere nuevas escalas de legitimidad, precisamente porque

la desinformación y la fragmentación del espacio público socavan los consensos democráticos. Cerny (2010) describe este proceso como neopluralismo transnacional, en el que las dinámicas globales reconfiguran las formas de dominación y resistencia, evidenciando que la democracia ya no puede concebirse exclusivamente desde el marco del Estado-nación. La ética del conocimiento, en este sentido, se vincula inseparablemente con la política de la voz: reconocer, acoger y responsabilizarse de la pluralidad de testimonios es condición para que el saber se convierta en un bien común y no en un instrumento de dominación y destrucción.

### **Continuidades y rupturas en la historia del pensamiento político**

El itinerario del pensamiento político se ha forjado en medio de tensiones fundacionales y desplazamientos conceptuales que, lejos de seguir una trayectoria lineal, ponen de relieve las continuidades y rupturas en la manera de concebir al sujeto social y sus formas de organización. El tránsito no ha sido meramente institucional, sino también epistemológico y simbólico. Como señala Lazzarich (2024), la reconstrucción de las genealogías políticas exige reconocer que los métodos y las epistemologías no son neutrales, sino que responden a contextos de poder y a disputas sobre la legitimidad del saber. En este sentido, la Polis no puede entenderse únicamente como un espacio originario de deliberación, sino como el punto de partida de una tensión

que se prolonga hasta la modernidad: la relación entre comunidad y autoridad, entre participación y exclusión.

La revisión de Skinner (2020) sobre los fundamentos del pensamiento político moderno permite advertir cómo las categorías de soberanía y representación emergen como respuestas a las crisis de legitimidad que acompañaron la transición hacia el Estado. Sin embargo, estas categorías no clausuran la pluralidad de interpretaciones, sino que la intensifican. Paniagua y Vergara (2022) muestra que la tradición republicana, lejos de ser homogénea, se bifurca entre lecturas elitistas y plebeyas de la constitución mixta, revelando que la tensión entre inclusión y exclusión atraviesa incluso los modelos que se presentan como universales. La continuidad, entonces, se expresa en la persistencia de la pregunta por quién tiene derecho a decidir, mientras que la ruptura se manifiesta en la multiplicación de formas de legitimidad.

La modernidad política, como advierte Urbinati (2014), se desplaza de la democracia de partidos hacia el plebiscito de la audiencia, transformando la relación entre ciudadanía y representación. Este tránsito no supone la desaparición de las mediaciones, sino su reconfiguración en un escenario donde la opinión pública se convierte en un actor central. Aquí se enlaza la reflexión de Rosanvallon, recuperada por Diehl (2019), sobre la importancia de los conceptos de representación y pueblo para comprender el



populismo contemporáneo: la continuidad radica en la necesidad de figuras que encarnen lo colectivo, mientras que la ruptura se evidencia en la emergencia de liderazgos que apelan directamente a la audiencia sin mediaciones institucionales.

En este marco, las ideologías renovadas y las formas comunales se presentan como respuestas críticas a las limitaciones del Estado moderno. Thomassen (2016), al revisar la obra de Laclau y Mouffe, subraya que la hegemonía no es un estado fijo, sino una construcción contingente que articula demandas diversas en torno a significantes comunes. Esta perspectiva dialoga con las epistemologías del Sur de Sousa Santos (2018), quien denuncia el epistemicidio y reivindica la necesidad de formas comunales de conocimiento que restituyan la voz de los sujetos históricamente marginados. La ruptura aquí no es solo conceptual, sino también epistémica: se trata de cuestionar la pretensión universal de las categorías modernas y abrir espacio a saberes situados. La utilización de categorías como la 'injusticia epistémica' y el 'testimonio' permite medir el alcance real de la posverdad en la subjetividad. No se trata solo de un cambio en la narrativa política, sino de una desautorización del conocimiento del sujeto. Al rescatar estos saberes desde las periferias, las epistemologías del sur se revelan no como un añadido temático, sino como la praxis de resistencia necesaria para reconstruir la agencia política frente a la desarticulación del simulacro.

La neutralidad y la pluralidad, como itinerario universal del sujeto social, constituyen quizá la tensión más persistente en la historia del pensamiento político. Benhabib, en la lectura de Valdivielso Navarro (2018), advierte que la ciudadanía cosmopolita se enfrenta al dilema entre conflicto y diferencia, mostrando que la neutralidad nunca es absoluta, sino que se construye en medio de disputas. Taylor (2004) complementa esta visión al señalar que los imaginarios sociales modernos configuran horizontes de sentido compartidos, pero siempre atravesados por la pluralidad de experiencias. Young (2000), por su parte, plantea que la democracia debe ir más allá de la deliberación, reconociendo al otro como condición de posibilidad de la libertad. En esta línea, Honneth (2001) y Kaulino (2010) insisten en que el reconocimiento es el fundamento de la libertad social, lo que implica que la neutralidad no puede desligarse de la pluralidad de voces que constituyen al sujeto.

Las reflexiones de Habermas (1992) sobre la relación entre hechos y normas, así como las de Honig (2009) sobre la política de la emergencia, muestran que la neutralidad institucional se ve constantemente desafiada por la irrupción de lo inesperado. Agamben (2010), al problematizar la figura del *homo sacer*, evidencia que la modernidad política se funda también en exclusiones radicales que ponen en cuestión la universalidad proclamada. Así, la continuidad se expresa en la búsqueda de marcos normativos que sostengan la convivencia, mientras que la ruptura se manifiesta en la

constatación de que dichos marcos nunca logran neutralizar del todo la pluralidad de experiencias y conflictos.

## **Conclusiones**

El recorrido por los itinerarios del pensamiento político realizado en este capítulo permite concluir que la configuración del sujeto social ha sido, y sigue siendo, el resultado de una disputa incesante por los regímenes de verdad. Desde la transparencia de la *parresía* en la polis griega hasta la consolidación de la razón institucional en el Estado moderno, el ser humano ha buscado en la verdad un suelo común sobre el cual edificar la convivencia. Sin embargo, el análisis del escenario contemporáneo revela que nos encontramos ante un punto de inflexión crítico: la posverdad no actúa meramente como una distorsión de los hechos, sino como una estrategia hegemónica de fragmentación que amenaza con disolver la capacidad de acción colectiva. Al erosionar la distinción entre el dato factual y la interpretación interesada, este fenómeno desarticula la agencia del sujeto, convirtiendo el espacio público en un campo de fuerzas donde el diálogo es sustituido por el antagonismo emocional y el aislamiento en burbujas informativas.

La integración de una narrativa plural nos ha permitido identificar que esta crisis es multidimensional. A través del lente de la epistemología social, se hace evidente

que la posverdad genera una forma de injusticia epistémica que silencia al sujeto al desautorizar su capacidad de conocer y testimoniar la realidad. Cuando la "infocracia" digital prioriza la eficacia del impacto sobre la veracidad del contenido, el ciudadano queda reducido a un consumidor de simulacros, perdiendo la potencia transformadora que una vez caracterizó su rol en la democracia. Frente a este diagnóstico, el capítulo sostiene que la respuesta no reside en un retorno a una razón tecnocrática y excluyente, sino en la apertura hacia las epistemologías del sur. Estas no representan un simple añadido temático, sino una praxis de resistencia que propone sanar el tejido social mediante el reconocimiento de los saberes situados y la recuperación de la justicia en el saber.

En última instancia, el aporte de esta investigación subraya que la reconstrucción del lazo político en el siglo XXI depende de nuestra capacidad para rescatar un modelo plural y emancipador del sujeto social. La superación de la posverdad exige un nuevo compromiso con la verdad pública, entendida no como un dogma estático, sino como un proceso deliberativo y plural. Recuperar la agencia del sujeto social implica, por tanto, transitar desde la fragmentación impuesta por el simulacro hacia una subjetividad capaz de reconocerse en la alteridad y en la responsabilidad compartida.

Solo a través de este giro hacia la justicia epistémica y el reconocimiento mutuo será posible rehabilitar el espacio

político como un lugar de encuentro, donde la palabra vuelva a tener la potencia de construir mundos más dignos, soberanos y profundamente humanos, capaz de articular justicia material y epistémica, participación democrática y reconocimiento de la diversidad, en un horizonte que entienda la política no como clausura, sino como práctica viva de igualdad, deliberación y responsabilidad colectiva. Se trata, en última instancia, de una apuesta por la dignidad humana, por la construcción de lo común y por la posibilidad de que la política recupere su condición originaria: ser, como afirmaba Aristóteles (2000), la actividad más alta del alma en consonancia con la virtud.

## Referencias

- Aaslund, A. (2024). *How can experiential knowledge contribute to epistemic justice? Dissenting voices and subaltern speech in social work* [Seminario]. Universidad de Gante, Bélgica.
- Adkins, A. (2017). Pathologies of testimony. En J. Lackey & E. Sosa (Eds.), *The epistemology of testimony* (pp. 253–271). Oxford University Press.
- Agamben, G. (2010). *Homo sacer: El poder soberano y la nuda vida*. Adriana Hidalgo Editora.
- Albareda, L., & Sison, A. J. G. (2020). Commons organizing: Embedding common good and institutions for collective action. Insights from ethics and economics.

- Journal of Business Ethics*, 166(4), 727–743.  
<https://doi.org/10.1007/s10551-020-04535-3>
- Aristóteles. (2000). *Política* (M. A. García Gual, Trad.). Editorial Gredos.
- Bertolotti, G. (2014). *Rumores, chismes y secretos: Una aproximación sociológica*. CLACSO.
- Bidaseca, K. (2011). Mujeres blancas buscando salvar a mujeres de color café. *Andamios. Revista de Investigación Social*, 8(17), 61–89.
- Briceño, H., & Maingon, T. (2015). *Avances y retrocesos de la democracia participativa en Venezuela*. Friedrich-Ebert-Stiftung.
- Butler, J. (2004). *Undoing gender*. Routledge.
- Byron, W. J. (2009). *Framing the principles of Catholic social thought*. Loyola Marymount University.
- Caldera, R. (1996). *Compromiso solidario* (Tomo III, Vols. I y II). Ediciones de la Presidencia de la República.
- Calla, P. (2010). Feminismos indígenas en América Latina: Ciudadanía, complementariedad y memoria ancestral. En R. A. Hernández Castillo (Ed.), *Mujeres indígenas en América Latina: Teoría y acción política* (pp. 45–68). CIESAS.
- Ceriani Cernadas, C. (2017). Rumores, chismes y secretos en la producción social de lo verosímil. *Apuntes de Investigación del CECYP*, (29), 146–155.
- Cerny, P. G. (2010). *Rethinking world politics: A theory of transnational neopluralism*. Routledge.

- Chávez Frías, H. (2005, 30 de enero). *Discurso en el V Foro Social Mundial*. Porto Alegre, Brasil.  
<https://rebellion.org/docs/10977.pdf>
- Chitty, N. (2008). *Consideraciones sobre política, economía y constitucionalismo*. Editorial Gráficas León.
- Cohen, N., & Garasic, M. D. (2024). Informed ignorance as a form of epistemic injustice. *Philosophies*, 9(3), 59.  
<https://doi.org/10.3390/philosophies9030059>
- Darder, A. (2017). Decolonizing interpretive research: Subaltern sensibilities and the politics of voice. *Qualitative Research Journal*, 18(2), 105–118.
- Daviault, P., & Vinay, J.-P. (1959). Le congrès de Bad Godesberg. *Journal des traducteurs / Translators' Journal*, 4(4), 147–153.
- Delannoy, S., & Mack, A. (2008). *El pensamiento político de Platón y Aristóteles*. Universidad Nacional de Rosario.
- de Sousa Santos, B. (2011). Epistemologías del Sur. *Utopía y Praxis Latinoamericana*, 16(54), 17–39.
- de Sousa Santos, B. (2014). *Epistemologies of the South: Justice against epistemicide*. Routledge.
- de Sousa Santos, B. (2016). *Epistemologies of the South*. UNESCO Chair on Community-Based Research and Social Responsibility.
- de Sousa Santos, B. (2018). *The end of the cognitive empire: The coming of age of epistemologies of the South*. Duke University Press.

- Diehl, P. (2019). Political theory through history. En O. Flügel-Martinsen (Ed.), *Pierre Rosanvallon's political thought: Interdisciplinary approaches* (pp. 39–60). Bielefeld University Press.
- Dieterich, H. (2007). *El socialismo del siglo XXI*. Fundación para la Investigación y la Cultura.
- Escobar, A. (2016). Thinking-feeling with the Earth: Territorial struggles and the ontological dimension of the epistemologies of the South. *Revista de Antropología Iberoamericana*, 11(1), 11–32.
- Federici, S. (2012). *Revolución en punto cero: Trabajo doméstico, reproducción y luchas feministas*. Traficantes de Sueños.
- Fraser, N. (2008). *Scales of justice: Reimagining political space in a globalizing world*. Columbia University Press.
- Fuentes Santibáñez, P. (2012). Algunas consideraciones en torno a la condición de la mujer en la Grecia antigua. *Intus-Legere Historia*, 6(1), 7–18.
- Garzón Rodríguez, C., & Acosta Navas, D. (2022). Injusticia epistémica: Una nueva epistemología para una antigua injusticia. *Estudios de Filosofía*, (66), 5–8.
- Giddens, A. (1999). *La tercera vía: La renovación de la socialdemocracia*. Taurus.
- Gómez Díaz de León, C. (2018). De la democracia participativa al gobierno abierto: Hacia una



- delimitación conceptual. *Política, Globalidad y Ciudadanía*, 4(7), 1–22.
- Gómez Peralta, H. (2012). Precisiones conceptuales sobre la democracia cristiana y el neoliberalismo. *Estudios Políticos*, 9(27), 121–133.
- Gough, I. (2015). *Welfare states and environmental states: A comparative analysis*. LSE Research Online.
- Groot, B., & Abma, T. (2022). Ethics framework for citizen science and public and patient participation in research. *BMC Medical Ethics*, 23(1), 1–9.
- Groot, B. C., Vink, M., Haveman, A., Huberts, M., Schout, G., & Abma, T. A. (2019). Ethics of care in participatory health research: Mutual responsibility in collaboration with co-researchers. *Educational Action Research*, 27(2), 286–302.
- Grosfoguel, R. (2011). Decolonizing post-colonial studies and paradigms of political economy. *Transmodernity Journal*, 1(1).
- Habermas, J. (1992). *Facticity and validity: Contributions to discourse theory of law and democracy*. MIT Press.
- Held, D. (2006). *Models of democracy* (3.<sup>a</sup> ed.). Polity Press.
- Hemerijck, A., Ronchi, S., & Plavgo, I. (2023). Social investment as a conceptual framework for analysing well-being returns and reforms in 21st century welfare states. *Socio-Economic Review*, 21(1), 479–500.
- Honneth, A. (2001). *La lucha por el reconocimiento*. Crítica.

- Honig, B. (2009). *Emergency politics: Paradox, law, democracy*. Princeton University Press.
- Jessop, B. (2015). *The state: Past, present, future*. Polity Press.
- Juan Pablo II. (1987, 30 de diciembre). *Sollicitudo rei socialis* [Carta encíclica]. Librería Editrice Vaticana.
- Kaulino, D. (2010). *Reconocimiento y libertad social: Una lectura de Axel Honneth*. Plural Editores.
- Knoll, M. (2017). Aristóteles y el pensamiento político aristocrático. *Revista de Filosofía*, 73, 87–106.
- Landau, M. (2004). *La ciudadanía como proceso histórico de construcción y reconstrucción*. Instituto de Investigaciones Gino Germani.
- Lander, E. (2014). *Crisis civilizatoria: Experiencias de los gobiernos progresistas y debates en la izquierda latinoamericana*. CALAS.
- Lazzarich, D. (2024). Genealogías políticas y estatuto epistemológico de la historia del pensamiento político. *Politics. Rivista di Studi Politici*, 22(2).
- Loewe, M., El-Haddad, A., Furness, M., Houdret, A., & Zintl, T. (2024). Drivers of change in social contracts: Building a conceptual framework. *Mediterranean Politics*.
- López Hana, S. (2014). Los aportes de Aristóteles a la ciencia política. *Procesos Históricos*, (25), 61–72.
- Marshall, T. H. (2005). *Citizenship and social class*. Pluto Press.

- Martínez, J. (1988). *El Príncipe de Maquiavelo y la ciencia política*. Tecnos.
- Meneses, M. P., & Bidaseca, K. (Coords.). (2018). *Epistemologías del Sur*. CLACSO.
- Mignolo, W. (2007). *La idea de América Latina: La herida colonial y la opción decolonial*. Gedisa.
- Mohanty, C. T. (2003). *Feminism without borders: Decolonizing theory, practicing solidarity*. Duke University Press.
- Mouffe, C. (1999). Deliberative democracy or agonistic pluralism? *Social Research*, 66(3), 745–758.
- Mouffe, C. (2005). *On the political*. Routledge.
- Nussbaum, M. C. (2000). *Women and human development: The capabilities approach*. Cambridge University Press.
- Paniagua, P., & Vergara, Á. (2022). Gobernanza policéntrica y la crisis de la democracia liberal. *Estudios Públicos*, 167, 77–105.
- Papachashvili, N., Roblek, V., Meško, M., & Podbregar, I. (2023). Welfare state on the theoretical crossroads: Analysis of the twenty-first-century studies. *International Review of Sociology*, 33(1), 23–38.
- Paredes, J. (2015). *Hilando fino desde el feminismo comunitario*. Asamblea Feminista Comunitaria.
- Pazos-Vidal, S. (2019). Subsidiarity and EU multilevel governance: Actors, networks and agendas. *Regional & Federal Studies*.

- Pérez C., M. (1998). *Glosario de términos de ciencia política*. Universidad Central de Venezuela.
- Quijano, A. (2014). Colonialidad del poder y clasificación social. En *Cuestiones y horizontes*. CLACSO.
- Quiroz Pacheco, Y. R., Chávez Cornejo, M. A., Gonzáles Flores, Y. N., Aguado Ortiz, J. J. A., & Marquina Ruiz, N. L. (2023). Las fake news como fenómeno de desinformación y posverdad. *Latam: Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4(6), 1474–1488.
- Quispe Arroyo, A. (2019). ¿La política es arte o ciencia de gobernar? *PURIQ*, 1(2), 253–260.
- Rancière, J. (2006). *Hatred of democracy*. Verso.
- Rodríguez Casallas, D. F., Páez Moreno, Á. E., Román Acosta, D., & Rodríguez Torres, E. (2024). Participación ciudadana, gobernanza democrática y derecho al desarrollo: Una revisión sistemática. *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 26(1), 198–214.
- Rosillo Olmedo, J. (2020). *La República de Platón: Justicia y política* [Trabajo de fin de grado, Universidad Pontificia Comillas].
- Saavedra Utman, J., & Santos, M. (2024). The public sphere and the Internet in Latin America. *Javnost: The Public*, 31(4), 459–476.
- Salmon, J. H. M. (1996). The legacy of Jean Bodin. *History of Political Thought*, 17(4), 500–522.

- Sasan, J. M. V. (2021). The social contract theories of Thomas Hobbes and John Locke. *Shanlax International Journal of Arts, Science and Humanities*.
- Sassoon, D. (1996). *One hundred years of socialism: The West European left in the twentieth century*. HarperCollins.
- Scheck, T. P. (2021). Thomas More: First and best apologist for Erasmus. *Moreana*, 58(1), 75–111.
- Skinner, Q. (2020). *Los fundamentos del pensamiento político moderno*. Fondo de Cultura Económica.
- Stoessel, S. (2014). Giro a la izquierda en la América Latina del siglo XXI. *Polis*, 13(39).
- Streeck, W. (2011). The crises of democratic capitalism. *New Left Review*, (71), 5–29.
- Taylor, C. (2004). *Modern social imaginaries*. Duke University Press.
- Thomassen, B. (2016). Hegemony and the politics of contingency: Laclau and Mouffe revisited. *Journal of Political Power*, 12(2), 224–241.
- Urbinati, N. (2014). *Democracy disfigured: Opinion, truth, and the people*. Harvard University Press.
- Valdivielso Navarro, J. (2018). La Europa cosmopolita de Seyla Benhabib. *Lectora*, 24, 111–126.
- Walsh, C. (Ed.). (2013). *Pedagogías decoloniales: Prácticas insurgentes de resistir, (re)existir y (re)vivir* (Tomo I). Abya-Yala.

- Wilde, M. (2015). *Contested spaces: The communal councils and participatory democracy in Chávez's Venezuela*. LSE Research Online.
- Wolin, S. S. (1953). Calvin and the Reformation. *The American Political Science Review*, 47(2), 524–540.
- Yépez, Á. (1988). *Historia universal*. Editorial Larense.
- Young, I. M. (2000). *Inclusion and democracy*. Oxford University Press.
- Zhang, Y. (2024). An analysis of social contract theory: Based on a comparative analysis of Hobbes, Locke and Rousseau. *Journal of Sociology and Ethnology*.

# **Cristianismo y democracia moderna: la influencia de la doctrina social de la Iglesia en la construcción de instituciones democráticas**

*Christianity and Modern Democracy: the Influence of the Social Doctrine of the Church on the Construction of Democratic Institutions*

**Javier Fattah<sup>57</sup>, Carolina Lagos Oróstica<sup>58</sup>**

## **Resumen**

El objetivo de este artículo es exponer que la democracia moderna y sus instituciones se han nutrido del pensamiento social y político de la doctrina cristiana.

---

<sup>57</sup> Universidad Católica de la Santísima Concepción.  
<https://orcid.org/0000-0001-8186-2734>

<sup>58</sup> Universidad Católica de la Santísima Concepción.  
<https://orcid.org/0000-0001-7858-2985>

Particularmente, desde los documentos eclesiásticos que expresan el compromiso de la Iglesia con el bien común y la dignidad de la persona. En primer lugar, se establece una descripción de la relación entre el cristianismo y la democracia moderna ilustrada. En segundo término, se identifica en la documentación eclesiástica la contribución del cristianismo a la democracia contemporánea, considerando especialmente la encíclica *Rerum Novarum* como declaración que establece una apertura hacia los derechos democráticos y la libertad religiosa. Para el cumplimiento de este propósito se emplea una metodología interpretativa de los documentos seleccionados, que permite transitar desde la inicial tensión entre cristianismo y modernidad, hacia la relación imprescindible entre valores cristianos y valores de la democracia.

**Palabras clave:** Cristianismo; democracia; dignidad humana; sociedad; política.

## **Abstract**

The purpose of this article is to show that modern democracy and its institutions have been nourished by the social and political thought of Christian doctrine. In particular, this has come from ecclesiastical documents that express the Church's commitment to the common good and the dignity of the person. First, a description of the relationship between Christianity and modern enlightened



democracy is established. Secondly, it identifies Christianity's contribution to contemporary democracy in ecclesiastical documentation, with particular reference to the encyclical *Rerum Novarum* as a declaration that establishes an openness towards democratic rights and religious freedom. To achieve this purpose, an interpretative methodology is used for the selected documents, which allows us to move from the initial tension between Christianity and modernity to the essential relationship between Christian values and the values of democracy.

**Keywords:** Christianity; democracy; human dignity; society; politics.

## **Introducción**

Tradicionalmente, desde la perspectiva cristiana, es posible distinguir dos modelos de democracia: la clásica y la moderna. Esta diferenciación no se limita a un aspecto histórico, sino que también se vincula con la fundamentación antropológica y ética que sostiene la idea misma de democracia (San Miguel Pérez, 2010). En su sentido clásico, la democracia se concibe como una herencia del pensamiento aristotélico y tomista, el cual establece que se trata de una forma de gobierno orientada al bien común de la sociedad. Dicho bien común se alcanza mediante la actuación virtuosa de los ciudadanos, capaces de reconocer un orden moral objetivo anterior a cualquier decisión política. Así, el bien

común no se reduce a la mera suma de intereses individuales, sino que constituye un bien compartido que favorece a todos los miembros de la comunidad. Según la definición clásica de Tomás de Aquino, el bien común es aquello que resulta bueno para cada uno y para todos los miembros de la comunidad (Baños Ardavín, 2017). Por ello, este principio implica necesariamente la cooperación social, la distribución justa de los recursos y la creación de condiciones políticas, sociales y económicas que permitan el desarrollo pleno de la persona.

Por otra parte, en la concepción moderna, la democracia se erige sobre las bases del contractualismo, especialmente en las formulaciones de Thomas Hobbes y John Locke (Sabine, 1968), quienes sostienen que el pacto social constituye el principio de legitimidad política y jurídica. Este modelo se articula en torno a la noción de una voluntad general, entendida como el fundamento que otorga validez a las leyes y derechos de los ciudadanos, concebidos como sujetos capaces de autodeterminación<sup>59</sup>. Posteriormente, el siglo XIX marca un periodo en el que la Iglesia Católica desarrolla un análisis crítico frente a estas nociones modernas, objetando la desvinculación de la política respecto de la verdad objetiva. Tal cuestionamiento se prolonga en la reflexión contemporánea, como se observa en la encíclica *Caritas in Veritate* (2009), donde se afirma que la verdad

---

<sup>59</sup> Al mencionar la autodeterminación lo hacemos en un sentido moderno y especialmente kantiano.

permite a los seres humanos ir más allá de las determinaciones culturales y las sensaciones subjetivas. En este sentido, la Iglesia plantea que la democracia no puede reducirse a la mera expresión de la voluntad, sino que requiere de un horizonte normativo sustentado en la verdad con la finalidad de orientar la acción política hacia el bien común.

La verdad abre y une el intelecto de los seres humanos en el logos del amor: éste es el anuncio y el testimonio cristiano de la caridad. En el contexto social y cultural actual, en el que está difundida la tendencia a relativizar lo verdadero, vivir la caridad en la verdad lleva a comprender que la adhesión a los valores del cristianismo no es sólo un elemento útil, sino indispensable para la construcción de una buena sociedad y un verdadero desarrollo humano integral. (Benedicto XVI, 2009)

En este sentido, la doctrina católica no sólo se interesa por la forma de gobierno democrático, también se propone ser un aporte para la sociedad occidental secular en su legítima aspiración de fortalecer los valores democráticos (Michellini & Figueroa, 2020). El reconocimiento de la dignidad humana, la justicia social y el bien común no son simples agregados a la sociedad; más bien, son contenidos que requieren de un sustento procedente de principios rectores no sencillamente convencionales. A este respecto, el objetivo de este texto es

exponer que la democracia moderna y sus instituciones se han nutrido histórica y culturalmente de un sustrato filosófico metafísico procedente del pensamiento de la doctrina cristiana.

Para el cumplimiento de este propósito se emplea una metodología interpretativa que permite centrar el análisis en los aspectos esenciales de las declaraciones eclesíásticas de los siglos XIX y XX. En este contexto, el artículo observa la convergencia entre la filosofía política de autores como Rawls, Habermas y Taylor, con la teoría democrática de pensadores como Maritain y Mounier, junto a la Doctrina Social de la Iglesia. Desde esta perspectiva pluralista, se propone una lectura de los documentos eclesíásticos que no se limita a una interpretación interna de la tradición cristiana, sino que busca establecer un diálogo con los debates actuales sobre legitimidad democrática y derechos humanos.

La tesis central de este artículo es de carácter genealógico-normativo, indicando que la democracia moderna se ha nutrido de un sustrato antropológico y ético de origen cristiano. Dicho sustrato sigue siendo fundamental para evitar que la democracia se reduzca a un mero procedimiento formal, desvinculado de la dignidad humana que le da sentido. Por lo tanto, como tesis secundaria se sostiene que la dignidad humana posee un fundamento metafísico previo a cualquier ordenamiento jurídico. Este fundamento, arraigado en la tradición cristiana, otorga a la

democracia un horizonte normativo que trasciende lo meramente institucional. La necesidad que se plantea en este trabajo es, por tanto, de orden cultural, filosófico e histórico. No se trata de una dependencia institucional entre Iglesia y Estado, sino de reconocer que los valores cristianos han configurado un trasfondo indispensable para la legitimidad democrática y la protección de los derechos humanos.

El orden de esta exposición se distribuye en una introducción, el desarrollo de cuatro puntos aclaratorios de los documentos seleccionados, y una conclusión reflexiva respecto a la relevancia de los valores cristianos en el desenvolvimiento de la democracia actual.

## **Cristianismo y democracia moderna**

La relación entre el cristianismo y la democracia ha sido un tema central para el pensamiento político moderno y contemporáneo. Acepciones concernientes al reconocimiento de la dignidad humana como derecho inalienable de toda persona, la exigencia ética de la justicia social, el principio de subsidiariedad y de solidaridad, han sido de hondo interés para la Iglesia y su doctrina, como también para la filosofía. Al respecto, la Doctrina Social de la Iglesia manifiesta la necesidad de promover la dignidad humana en la época moderna.

The movement towards the identification and proclamation of human rights is one of the most significant attempts to respond effectively to the inescapable demands of human dignity. The Church sees in these rights the extraordinary opportunity that our modern times offer, through the affirmation of these rights, for more effectively recognizing human dignity and universally promoting it as a characteristic inscribed by God the Creator in his creature. (Pontifical Council for Justice and Peace, 2004, p. 152)<sup>60</sup>

La afirmación citada conduce a reafirmar la centralidad de la dignidad humana dentro del pensamiento cristiano y la cultura occidental. De acuerdo con la tradición Tomista y Agustina, la dignidad humana no es una convención social o un contrato, ante todo es una propiedad inherente del ser humano como criatura creada a imagen de Dios. En este sentido, los derechos humanos y, por ende, los sistemas democráticos que han buscado su realización mediante la jurisdicción y la actividad política poseen un antecedente metafísico que se remonta a un antecedente

---

<sup>60</sup> “El movimiento hacia la identificación y proclamación de los derechos humanos es uno de los intentos más significativos para responder eficazmente a las exigencias ineludibles de la dignidad humana. La Iglesia ve en estos derechos la extraordinaria oportunidad que nos ofrece nuestra época moderna, a través de la afirmación de estos derechos, para reconocer más eficazmente la dignidad humana y promoverla universalmente como una característica inscrita por Dios Creador en su criatura”. (Traducción propia).

metafísico de la teología cristiana (Maritain, 1943). Esta idea fue desarrollada con profundidad por Jacques Maritain, quien sostiene que la dignidad humana constituye el punto de partida de cualquier orden político justo. Para este autor, la democracia moderna no surge únicamente de acuerdos históricos o de prácticas institucionales, sino de un reconocimiento de la persona como ser trascendente, dotado de valor absoluto y orientado hacia el bien. De este modo, la tradición cristiana aporta un fundamento normativo que trasciende lo meramente procedimental y asegura que la democracia se mantenga vinculada a la dignidad de la persona como núcleo irrenunciable de su legitimidad.

No obstante, desde la modernidad ilustrada, la relación entre cristianismo y democracia moderna no ha estado exenta de tensión. Para comprender este fenómeno, es necesario situarse en el hecho que la democracia moderna es heredera del racionalismo del siglo XVIII y de la Revolución Francesa. Ambos momentos fueron posibles gracias a la separación de la ontología clásica y concepción del ser humano como criatura racional. El pensamiento ilustrado mantiene la idea secular de una razón autónoma que orienta las actuaciones humanas. Por consiguiente, la vinculación entre orden político y Ley de Dios se desarticula, provocándose un proceso político y social de transferencia de los conceptos teológicos a la teoría moderna del Estado. Al producirse el reemplazo de la omnipotencia divina por la preponderancia del legislador,

la autoridad política reside en la soberanía popular<sup>61</sup> y la voluntad general.<sup>62</sup>

De acuerdo con Rafael Gambra (1982), el desmantelamiento de la sociedad medieval se produjo por una negación completa “de un orden político en consonancia con la Ley de Dios y el orden natural dispuesto por Él” (p. 94). Esta ruptura, expresada en la Declaración de los Derechos del Hombre y del Ciudadano (1789), introdujo una concepción secular del sujeto, cuyos derechos legítimos se derivan de un contrato social sin referencia trascendente. Sin embargo, esta transformación no eliminó del todo el sustrato religioso, que permaneció en el plano simbólico de la política como una base cultural y espiritual latente. En este punto, la reflexión de Jacques Maritain resulta esencial, pues afirma que la dignidad humana constituye el fundamento metafísico previo a cualquier ordenamiento jurídico y el núcleo irrenunciable de un sistema democrático justo. Así, mientras la modernidad política intentó desligarse de lo sagrado, la tradición cristiana mantuvo vigente la idea de que la democracia no puede

---

<sup>61</sup> En sentido moderno, se puede entender como el principio según el cual todo poder político proviene del pueblo, entendido como la fuente última de la autoridad.

<sup>62</sup> Concepto planteado por Rousseau que considera la voluntad colectiva orientada al bien común, distinta de la suma de intereses particulares. En la teoría moderna, es el criterio que legitima las leyes: una norma es justa cuando expresa esta voluntad general, es decir, aquello que la comunidad reconoce como conveniente para todos.



reducirse a un mero procedimiento formal, sino que requiere reconocer a la persona como ser trascendente, dotado de valor absoluto y orientado hacia el bien.

Cabe reconocer que tradiciones no confesionales también han desarrollado fundamentos sólidos para la democracia y los derechos humanos. El liberalismo político de John Rawls, por ejemplo, propone la noción de una razón pública compartida por ciudadanos con distintas concepciones del bien, sin necesidad de apelar a doctrinas religiosas (Rawls, 1995). De manera complementaria, el republicanismo cívico defendido por autores como Philip Pettit (1997) plantean una noción de libertad política basada en la no dominación y en la participación activa, sin requerir una antropología trascendente como punto de partida. Sin embargo, como ha mostrado Charles Taylor, la modernidad secular no implica necesariamente la extinción de lo religioso, sino más bien una reconfiguración de su lugar en la vida pública (Taylor, 2007). En la misma línea, Jürgen Habermas, en diálogo con Benedicto XVI, ha observado que las tradiciones religiosas conservan un “potencial semántico” relevante para las democracias contemporáneas (Habermas, 2006). La presente argumentación no pretende negar la validez de estos enfoques seculares, sino mostrar que el pensamiento cristiano ha aportado históricamente una dimensión antropológica complementaria: la afirmación de una dignidad humana que no se reduce a acuerdos

contingentes entre ciudadanos, sino que los precede y los fundamenta.

Por otra parte, para Carl Schmitt (2009), la teoría moderna del Estado no es otra cosa que una conceptualización secular de nociones teológicas. Esta noción sugiere que las categorías fundamentales de la política moderna —soberanía, autoridad, legitimidad y orden jurídico— no surgieron de manera autónoma, sino que fueron heredadas y transformadas a partir de conceptos originalmente teológicos. En otras palabras, la secularización no implicó una ruptura absoluta con la tradición religiosa, sino una traducción de sus nociones al lenguaje político moderno. Así, la idea de soberanía, por ejemplo, puede entenderse como una transposición del concepto teológico de la omnipotencia divina, mientras que la noción de orden jurídico remite a la idea de un cosmos regulado por leyes naturales y divinas. De este modo, Schmitt pone de relieve el papel filosófico del cristianismo en la génesis del pensamiento político moderno, mostrando que incluso en su forma secular, la teoría del Estado conserva huellas de su origen teológico. Al respecto, el autor sostiene:

All significant concepts of the modern theory of the state are secularized theological concepts not only because of their historical development - in which they were transferred from theology to the theory of the state, whereby, for example, the omnipotent God

became the omnipotent lawgiver....<sup>63</sup> (Schmitt, 2009, p. 36)

Ahora bien, escritos como la encíclica *Quanta Cura* (1864 de Pío IX) y *Syllabus Errorum*, la Iglesia expresa su oposición a los principios del liberalismo moderno, tales como la separación entre la Iglesia y el Estado y la autonomía de la razón frente a la fe, es decir, frente a la derivación de un programa cultural moderno completamente antropocéntrico. La preocupación de la Iglesia respecto a una visión del individuo y de la sociedad enteramente desvinculada de la verdad revelada y de la noción de ley natural, no se trata de una cuestión puramente especulativa, sino, de una problemática asumida desde su praxis social (Maritain, 2001; Witte, 1993). Seguidamente, a lo largo de los siglos XIX y XX, la examinación de la Iglesia se encamina a destacar el aporte del núcleo ético del cristianismo a la democracia moderna contemporánea. Así, por ejemplo, Félicité de Lamennais fue uno de los primeros pensadores que intentó reconciliar el Evangelio con las libertades modernas, indicando que la libertad política y civil es una tendencia natural de la sociedad.

---

<sup>63</sup> “Todos los conceptos significativos de la teoría moderna del Estado son conceptos teológicos secularizados, no solo por su desarrollo histórico — en el que pasaron de la teología a la teoría del Estado, por lo que, por ejemplo, el Dios omnipotente se convirtió en el legislador omnipotente...— “. (Traducción propia).

Le but vers lequel tend la société..., est la liberté religieuse, politique et civile, c'est-à-dire d'un côté l'affranchissement de l'intelligence plus ou moins asservie sous tous les gouvernements modernes à la force brute du pouvoir, et de l'autre une extension de la sphère d'activité publique et particulière proportionnée aux développements de cette même intelligence, avec les garanties nécessaires des droits résultant de ce nouvel état social.<sup>64</sup> (Vacant, 1908, p. 2489)

Aunque la tesis del intelectual francés fue reprobada en la encíclica *Mirari vos* (1832) de Gregorio XVI, constituye uno de los planteamientos iniciales del catolicismo liberal, proceso que encontraría una respuesta en el siglo XIX con la política del *Ralliement* impulsada por León XIII. A través de esta directriz, la autoridad papal instó a los católicos franceses a aceptar la República sin abandonar los principios eclesiásticos ni renunciar a la fe, mostrando así la capacidad del cristianismo de dialogar con nuevas formas políticas sin perder su fundamento filosófico cristiano. Este proceso

---

<sup>64</sup> “El objetivo al que aspira la sociedad... es la libertad religiosa, política y civil, es decir, por un lado, la liberación de la inteligencia, más o menos sometida en todos los gobiernos modernos a la fuerza bruta del poder, y, por otro, una ampliación del ámbito de la actividad pública y privada proporcional al desarrollo de esa misma inteligencia, con las garantías necesarias de los derechos que se derivan de este nuevo estado social”. (Traducción propia).

culminó con la encíclica *Rerum Novarum* (1891), que marcó un hito en la doctrina social de la Iglesia al reconocer la necesidad de integrar la justicia social y los derechos de los trabajadores en el marco democrático. Posteriormente, el Concilio Vaticano II consolidó esta evolución mediante documentos como *Dignitatis Humanae* (1965a) y *Gaudium et Spes* (1965b), donde se definió la aceptación de la libertad religiosa y la participación democrática como derechos inherentes a la persona, conformes a su naturaleza humana. De este modo, el cristianismo no solo ofreció resistencia inicial a la modernidad política, sino que terminó por aportar un fundamento filosófico y antropológico decisivo para la legitimidad de la democracia contemporánea.

### **La doctrina cristiana en la esfera pública**

Para los católicos adherentes al cristianismo social y a la filosofía personalista, la norma del *Ralliement* constituyó una oportunidad de apertura a la acción cristiana dentro del régimen republicano. En adelante, la situación de la clase trabajadora es asumida como un problema social que debe resolverse con medidas superiores a la caridad individual de cada católico. Al respecto, se puede mencionar lo indicado por Bryk (1993) en su comprensión de la acción social que emprendía el cristiano católico a mediados del siglo XX, cuando surge que la problematización de la justicia social:

Beginning in the mid-nineteenth century, however, some European lay Catholics and Individual clerics began to see the "working-class problem" as a structural issue of social justice that required more than individual charity. These "social Catholics" worked to develop a response that balanced the excesses of the two dominant ideologies of the day: economic liberalism, which overemphasized the individual and ignored the moral implication of the relation of capital to labor; and totalitarian socialism, which gave too much control to the state and devalued the dignity of the person.<sup>65</sup> (Bryk, 1993, p. 41)

En esta línea, la publicación de la encíclica *Rerum Novarum* de León XIII se presentó taxativamente opuesta al liberalismo económico y al marxismo. Tanto desde un punto de vista teórico, como desde una propuesta política inspirada en los principios cristianos, el documento de León XIII plasma una concepción social integral. Por lo tanto, la

---

<sup>65</sup> "Sin embargo, a partir de mediados del siglo XIX, algunos católicos laicos europeos y clérigos individuales comenzaron a considerar el «problema de la clase trabajadora» como una cuestión estructural de justicia social que requería algo más que la caridad individual. Estos «católicos sociales» trabajaron para desarrollar una respuesta que equilibrara los excesos de las dos ideologías dominantes de la época: el liberalismo económico, que daba demasiada importancia al individuo e ignoraba las implicaciones morales de la relación entre el capital y el trabajo; y el socialismo totalitario, que otorgaba demasiado control al Estado y devaluaba la dignidad de la persona". (Traducción propia).

encíclica como primera intervención sistemática del magisterio pontificio en cuestiones sociales y económicas, constituye una declaración cardinal de la Iglesia Católica, no sólo por influir en las personas católicas, sino, por entregar a la cultura occidental un referente para la conducta ética frente a la distribución y adquisición de bienes producidos por el trabajo de la persona y, en consecuencia, la retribución en el salario justo. A este respecto, el Papa Francisco ha señalado que los actuales problemas económicos, sociales y políticos son producto de una realidad estructural que los sustenta (Orellana & Johnson, 2023). En continuidad con esta reflexión, puede recordarse que la encíclica que aquí se analiza planteaba ya en su tiempo cuestiones novedosas frente a una economía que no obedecía a ningún principio moral, pues existía una libre competencia sin freno alguno, por lo que cualquiera pudiera sacar provecho personal a costa del sufrimiento ajeno (González, 2024). Esta crítica, formulada en el contexto de la industrialización del siglo XIX, conserva plena vigencia en el presente, donde las problemáticas derivadas de la desigualdad y la falta de regulación siguen teniendo gran presencia. De este modo, la doctrina cristiana se muestra no solo como un referente religioso, sino como un aporte cultural y político que insiste en la necesidad de vincular la economía y la vida pública con la dignidad humana como valor trascendente, evitando que la democracia y el orden social se reduzcan a meros mecanismos formales desprovistos de orientación ética.

Las formulaciones de la *Rerum Novarum*, como se ha indicado, albergan un carácter político que, junto con destacar el reconocimiento del trabajo humano como una realidad irreductible, debe cuidar que las leyes del Estado tengan fundamento la obtención del bien común. En este sentido, “esta encíclica funge como el documento oficial que regula y manifiesta explícitamente la relación entre la Iglesia y el Estado” (González, 2024, p. 6). De ahí la comprensión de que la dignidad intrínseca del trabajador sea una condición que obliga a la política a generar condiciones laborales coherentes con el valor del ser humano. Además, la propiedad privada es reconocida como una expresión del orden natural, por tanto, la *Rerum Novarum* encausa la deliberación de los gobiernos para atender la defensa de la dignidad del trabajador y su derecho a la propiedad. En este sentido, la Iglesia manifiesta la obligatoriedad de la cooperación mutua entre clases sociales con la finalidad de resguardar los derechos de los débiles y los pobres (Juan Pablo II, 1991).

Como se puede observar, la contribución cristiana a la democracia no se desarrolla al margen del constitucionalismo moderno, sino que lo nutre desde su interior. La consagración de los derechos fundamentales en las constituciones del siglo XX, la separación de poderes como garantía contra el abuso de autoridad y la protección de las minorías son instituciones que, aunque se consolidaron en un marco secular, encuentran su raíz en una tradición cristiana que siempre subordinó el poder político al bien de la persona. En este sentido, el aporte



del cristianismo no pretende reemplazar ni competir con las mediaciones jurídicas e institucionales propias de la democracia moderna, sino ofrecer el sustento antropológico que les da coherencia y legitimidad.

Hasta aquí, se puede sostener que la evolución del pensamiento católico político es proceso de integración de distintos elementos doctrinales que propician la generación de respuestas frente al pragmatismo que se instala en el poder y la acción política contemporánea. A juicio de Luis Caeiro (2022), gracias a los documentos sociales de la Iglesia, se manifiesta una posibilidad alternativa frente al capitalismo liberal y el colectivismo marxista. La forma que propone la entidad religiosa se trata de una acción política basada en la noción de pueblo que posee su unidad en la solidaridad y el compromiso con el bien común. Según declara el mencionado autor, las encíclicas:

They became spokespersons of an alternative or intermediate ‘third way’ between, or perhaps beyond, liberal individualism and unfettered capitalism, on the one hand, and socialist collectivism and Soviet-style communism on the other. This ‘way’ took shape as a distinctive political action, based on and nurtured by a constellation of political concepts and ideas that descended from Catholic philosophy and social teaching. These included, in addition to personalism, key concepts such as anti-materialism, popularism—

the idea of a ‘people’ whose unity come from within and from the solidarity between its members, all committed to the temporal common good—and subsidiarity—the principle of distribution of power downward to the local level, as close as possible to the citizens.<sup>66</sup> (Luis Caeiro, 2022, p. 11)

Posteriormente, a partir del Concilio Vaticano II surgen la *Dignitatis humanae* (1965) y *Gaudium et spes* (1965). Ambas declaraciones patentizan una actitud abierta hacia el pluralismo y la democracia. En *Dignitatis Humanae* el “problema” de la libertad religiosa se presenta como un derecho, estableciendo que: “Este Concilio Vaticano declara que la persona humana tiene derecho a la libertad religiosa” (Concilio Vaticano II, 1965).

El principio de libertad, aunque pueda parecer de una simpleza inicial, implica en realidad el reconocimiento de la

---

<sup>66</sup> “Se convirtieron en portavoces de una «tercera vía» alternativa o intermedia entre, o quizás más allá, del individualismo liberal y el capitalismo sin restricciones, por un lado, y el colectivismo socialista y el comunismo de estilo soviético, por otro. Esta «vía» tomó forma como una acción política distintiva, basada y alimentada por una constelación de conceptos e ideas políticas que descendían de la filosofía católica y la doctrina social. Entre ellos se encontraban, además del personalismo, conceptos clave como el antimaterialismo, el populismo —la idea de un «pueblo» cuya unidad proviene de dentro y de la solidaridad entre sus miembros, todos comprometidos con el bien común temporal— y la subsidiariedad —el principio de distribución del poder hacia abajo, al nivel local, lo más cerca posible de los ciudadanos—.”. (Traducción propia).

existencia de otras creencias, lo que abre la posibilidad de una alineación concreta entre la concepción cristiana y la democracia. En este sentido, la relación entre la doctrina de fe y la epistemología política confirma la compatibilidad de la fe con una forma de gobierno que surge en términos seculares y desde el racionalismo. A partir de ello, se sostiene que una política social carente de la guía de la conciencia cristiana carece de una recta eficacia y limita la armonía entre los ciudadanos. Por consiguiente, resulta imprescindible mantener la relación entre la Iglesia y el Estado, aun cuando se trate de entidades distintas (González, 2024), pues dicha vinculación asegura la orientación ética y espiritual de la vida democrática.

Dentro de esta línea, la *Gaudium et spes* visualiza la legitimidad de las distintas formas de gobierno, en tanto sean garantes de los derechos fundamentales del ser humano y no se desvíen de la búsqueda del bien común. Se advierte aquí, que la relación entre la Iglesia y la modernidad se funde en la preocupación de velar por las libertades ciudadanas por medio de leyes e instituciones. De manera que, sin instituir un modelo político específico de gobierno, el documento eclesiástico mantiene inviolable el reconocimiento de la dignidad humana que ha de ser partícipe de la vida pública que es, en definitiva, expresión del ser político del ser humano.

El Concilio Vaticano II, en *Gaudium et Spes*, explica que la naturaleza humana admite la constitución de estructuras políticas y jurídicas que permitan a los ciudadanos participar, sin discriminación, en el establecimiento de la comunidad política, en la gestión de la cosa pública, en la determinación de los límites de las instituciones y en la elección de los gobernantes. En este sentido, “la Iglesia alaba y estima la labor de quienes, al servicio del hombre, se consagran al bien de la cosa pública y aceptan las cargas de este oficio” (Concilio Vaticano II, 1965). Este reconocimiento subraya el valor cultural y político del cristianismo, que no se limita a la esfera religiosa.

En este marco, la democracia no puede entenderse únicamente como un entramado de instituciones y normas jurídicas, sino que se sostiene en una concepción más profunda de la persona humana concebida como un ser libre y responsable, llamado a participar activamente en la vida pública. Así, la doctrina cristiana se revela no solo como un legado religioso, sino como un aporte cultural decisivo en la configuración de un orden democrático que reconoce la libertad y la responsabilidad de cada individuo en la construcción del bien común.

## **La ciudadanía cristiana**

En el marco del cristianismo, la noción de ciudadanía va más allá de un concepto jurídico, político o cívico, como

puede suceder en las democracias. La ciudadanía no es únicamente una categoría que define derechos y deberes dentro de un Estado, más bien, la ciudadanía cristiana está vinculada con la noción de bien común, con la participación de la persona en la comunidad y con la trascendencia. Por lo tanto, comprender la ciudadanía desde el cristianismo invita a mirar más allá de los aspectos terrenales, para considerar la orientación hacia la trascendencia. En efecto, no considerar esta perspectiva, implica una percepción antropológica, ética y cultural limitada para el desarrollo de la vida ciudadana. De acuerdo con el Papa Francisco en la actual “cultura del descarte” (Orellana & Johnson, 2023) la orientación política y ciudadana no ubica en su centro a la persona humana en su dimensión trascendente, sino, únicamente desde las posibilidades de su acción útil. Al respecto, es posible sostener que la postulación cristiana de una dignidad humana intrínseca sí puede ser parte de la realización histórica de la democracia a través de la actuación moral de los pueblos. Al recoger la concepción antropológica cristiana, las democracias actuales pueden afirmar el valor insustituible del ser humano, así como incrementar en sus determinaciones civiles una ética de la solidaridad basada en el reconocimiento mutuo entre personas iguales.

Lo antes señalado introduce la importancia de contar con la libertad humana, pues, es imposible concebir la vida ciudadana sin libertad. El cristianismo sostiene que la libertad se realiza plenamente en cuanto el bien es objeto de

elección, es decir, la libertad no sería un acto aislado o motivado por una sensación del existir. La libertad es realizada mediante una dinámica relacional que involucra a la persona con la comunidad. Al respecto Benedicto XVI muestra en su *Caritas in Veritate* (2009), que la libertad no debe ser concebida en desunión de la verdad. Significa que la libertad no es simplemente la capacidad de elegir o la capacidad de la mayoría de determinar una decisión; más bien, la libertad posee la limitación de la verdad que no depende de factores contingentes, pero sí de principios universales de bien y justicia (Gewirth, 1978). A este respecto, se puede considerar que el cristianismo ha aportado a la democracia moderna la idea de que la libertad política y la legitimidad de las leyes no tienen base en la voluntad de la mayoría como sostiene, precisamente, la democracia moderna ilustrada. En efecto, sobre el derecho de la mayoría a decidir en las cuestiones públicas, el cristianismo agrega que tales decisiones no podrían justificarse si no son orientadas por principios éticos superiores descubiertos por la razón. De esta manera, el pensamiento cristiano apela a una verdad objetiva como fundamento de la democracia; por ende, no busca imponer una doctrina confesional a una sociedad pluralista. Es más, si seguimos la distinción tomista entre razón natural y revelación, los principios de dignidad humana, justicia y bien común son accesibles racionalmente, con independencia de la fe. En este sentido, la propuesta cristiana no es teológica sino filosófica, se busca que las

instituciones democráticas reconozcan que sus fundamentos normativos no son puramente convencionales, sino que responden a una realidad humana que precede a cualquier acuerdo político.

En otras palabras, la voluntad popular no puede ser la fuente absoluta de legitimidad sin ser regulada por una moralidad objetiva, enfocada en la dignidad de la persona humana. Por lo tanto, la democracia desde la doctrina cristiana es asumida en virtud de la participación de la verdadera libertad que se ejerce dentro de un marco ético que, como se ha venido sosteniendo, se orienta siempre al bien común y el cuidado de la dignidad de la persona. De esta manera, la voluntad de la mayoría estaría directamente vinculada la responsabilidad moral, pues, la libertad no es únicamente actual según la voluntad propia, más bien, es la posibilidad de vivir en una comunidad sostenida en los principios de justicia y solidaridad y de dirigirse al bien verdadero (Dewey, 1927; Sandel, 1998; Sevelsted & Toubøl, 2022).

## **Conclusión**

El recorrido a través de la historia y la doctrina política del cristianismo permite afirmar que la relación entre fe y democracia no debe entenderse como una cuestión insuperable, sino como un proceso de maduración que ha permitido mejorar los fundamentos éticos de la vida en

sociedad. Como se ha demostrado, la evolución del pensamiento católico político es el resultado de un proceso dialéctico que integra la fidelidad doctrinal con una respuesta ante la secularización. Desde las primeras condenas al liberalismo hasta la aceptación del pluralismo en el Concilio Vaticano II, la Iglesia no ha cambiado su núcleo de verdad, sino que ha profundizado en la comprensión de la dignidad humana presentando una tercera vía entre el individualismo liberal y las corrientes de carácter marxista. La democracia dentro del cristianismo no es un simple procedimiento formal; más bien, se traduce como un imperativo moral orientado hacia la protección de la persona y de su dignidad.

Un elemento esencial que se desprende del análisis es la distinción entre la democracia moderna, fundada en el contractualismo y la voluntad general, y la democracia clásica o cristiana, orientada al bien común y la verdad objetiva. Mientras que la tradición de Thomas Hobbes y John Locke sitúan la legitimidad en la autodeterminación colectiva sin referencia a un orden moral previo. Por su parte, el pensamiento cristiano considera que una democracia desvinculada de la verdad corre el riesgo de transformarse en una tiranía donde primen los intereses particulares. Así, Benedicto XVI, en *Caritas in Veritate*, pone de manifiesto este peligro al señalar que una libertad sin verdad no es auténtica libertad; la autonomía absoluta del individuo, entendida como la capacidad de elegir sin restricciones, puede terminar en autoritarismo si no se reconoce que la



libertad humana debe ejercerse dentro de unos límites orientadores en dirección al bien común y la verdad objetiva.

Una objeción frecuente a este planteamiento es que la apelación a principios morales objetivos en la política democrática puede derivar en los que podemos entender como un perfeccionismo coercitivo, es decir, el uso del poder del Estado para imponer una concepción particular del bien. Frente a esta objeción, cabe señalar que la doctrina social de la Iglesia señala expresamente la existencia de la libertad religiosa y de conciencia como derechos inalienables.

En este sentido, la relevante contribución del cristianismo a la democracia contemporánea y la cultura occidental es su concepción antropológica donde el ser humano no es un individuo aislado, sino una persona relacional. Además, al definir a la persona como imagen y semejanza de Dios, la antropología cristiana establece que los derechos humanos no son convenciones sociales ni concesiones del Estado, ante todo son propiedades inherentes a la naturaleza humana, verdades metafísicas previas a cualquier ordenamiento jurídico.

Lo anterior, implica que la ciudadanía cristiana va más allá del estatus legal; la ciudadanía es una categoría moral que vincula al individuo con la comunidad y con la trascendencia, exigiendo una participación en la vida pública que no se agota en el voto de elección, sino que busca la construcción de

condiciones justas para todos. Así, por ejemplo, Jacques Maritain (2021) afirma que la democracia es la única expresión política verdadera del cristianismo, en tanto que sólo en ella se pueden realizar plenamente los ideales evangélicos de fraternidad y dignidad.

## Referencias

- Asamblea Nacional Constituyente de Francia. (1789). *Declaración de los Derechos del Hombre y del Ciudadano*. Conseil constitutionnel.  
[https://www.conseil-constitutionnel.fr/sites/default/files/as/root/bank\\_m/m/espagnol/es\\_ddhc.pdf](https://www.conseil-constitutionnel.fr/sites/default/files/as/root/bank_m/m/espagnol/es_ddhc.pdf)
- Baños Ardavín, E. J. (2017). Aproximación a la noción de bien común en Tomás de Aquino. *Metafísica y Persona*, (12).  
<https://doi.org/10.24310/Metyper.2014.voi12.2729>
- Benedicto XVI. (2009). *Caritas in veritate* [Carta encíclica]. Libreria Editrice Vaticana.
- Bryk, A. S. (1993). *Catholic schools and the common good*. Harvard University Press.
- Concilio Vaticano II. (1965a). *Dignitatis humanae* [Declaración sobre la libertad religiosa]. Libreria Editrice Vaticana.  
[https://www.vatican.va/archive/hist\\_councils/ii\\_vatican\\_council/documents/vat-ii\\_decl\\_19651207\\_dignitatis-humanae\\_sp.html](https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decl_19651207_dignitatis-humanae_sp.html)

- Concilio Vaticano II. (1965b). *Gaudium et spes* [Constitución pastoral sobre la Iglesia en el mundo actual]. Libreria Editrice Vaticana.  
[https://www.vatican.va/archive/hist\\_councils/ii\\_vatican\\_council/documents/vat-ii\\_const\\_19651207\\_gaudium-et-spes\\_sp.html](https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19651207_gaudium-et-spes_sp.html)
- Dewey, J. (1927). *The public and its problems*. Henry Holt and Company.
- Gambra, A. (1982). *La Revolución Francesa: Mito y realidad*. Speiro.
- Gewirth, A. (1978). *Reason and morality*. University of Chicago Press.
- González, G. (2024). La concepción de política social del Papa León XIII a la luz de la encíclica *Rerum Novarum*. *Ponto-e-Vírgula*, 33(1).  
<https://doi.org/10.23925/1982-4807.2023i33e53702>
- Gregorio XVI. (1832). *Mirari vos* [Carta encíclica]. Libreria Editrice Vaticana.  
<https://www.vatican.va/content/gregorius-xvi/en/documents/encyclica-mirari-vos-15-augusti-1832.html>
- Habermas, J. (2006). Religion in the public sphere. *European Journal of Philosophy*, 14(1), 1–25.  
<https://doi.org/10.1111/j.1468-0378.2006.00241.x>
- Juan Pablo II. (1991). *Centesimus annus* [Carta encíclica]. Libreria Editrice Vaticana.

- León XIII. (1891). *Rerum novarum* [Carta encíclica]. Libreria Editrice Vaticana.  
[https://www.vatican.va/content/leo-xiii/es/encyclicals/documents/hf\\_l-xiii\\_enc\\_15051891\\_rerum-novarum.html](https://www.vatican.va/content/leo-xiii/es/encyclicals/documents/hf_l-xiii_enc_15051891_rerum-novarum.html)
- Luis Caeiro, P. B. (2022). The educational project of John Paul II and Benedict XVI: A philosophical and theological approach. *Religions*.
- Maritain, J. (1943). *Education at the crossroads*. Yale University Press.
- Maritain, J. (2001). *Los derechos del hombre y la ley natural. Cristianismo y democracia*. Ediciones Palabra.
- Maritain, J. (2021). *El hombre y el Estado*. Ediciones Olejnik.
- Michellini, D., & Figueroa, M. (2020). *Religión y espacio público: Estudios sobre J. Habermas y R. Rorty*. RIL.
- Mounier, E. (2011). *Personalism*. Read Books Ltd.
- Orellana, F., & Johnson, J. (2023). De la cultura del descarte a la cultura del encuentro: El aporte del Papa Francisco a la Doctrina Social de la Iglesia. *Anales de Teología*, 25(2), 293–324. <https://doi.org/10.21703/2735-634520232522242>
- Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Oxford University Press.
- Pío IX. (1864a). *Quanta cura* [Carta encíclica]. Libreria Editrice Vaticana.  
<https://www.vatican.va/content/pius->

[ix/it/documents/encyclica-quanta-cura-8-decembris-1864.html](https://www.vatican.va/content/pius-ix/it/documents/encyclica-quanta-cura-8-decembris-1864.html)

Pío IX. (1864b). *Syllabus complectens praecipuos nostrae aetatis errores* [Syllabus errorum]. Libreria Editrice Vaticana. <https://www.vatican.va/content/pius-ix/it/documents/encyclica-quanta-cura-8-decembris-1864.html>

Pontifical Council for Justice and Peace. (2004). *Compendium of the social doctrine of the Church*. Libreria Editrice Vaticana. [https://www.vatican.va/roman\\_curia/pontifical\\_councils/justpeace/documents/rc\\_pc\\_justpeace\\_doc\\_20060526\\_compendio-dott-soc\\_en.html](https://www.vatican.va/roman_curia/pontifical_councils/justpeace/documents/rc_pc_justpeace_doc_20060526_compendio-dott-soc_en.html)

Rawls, J. (1995). *Liberalismo político* (S. R. Madero Báez, Trad.). Fondo de Cultura Económica.

Sabine, G. H. (1968). *Historia de la teoría política*. Fondo de Cultura Económica.

San Miguel Pérez, E. (2010). *La democracia cristiana y la democracia de los cristianos*. Fundación Universitaria Española.

Sandel, M. J. (1998). *Liberalism and the limits of justice*. Cambridge University Press.

Schmitt, C. (2009). *Teología política: Cuatro capítulos sobre el concepto de soberanía*. Trotta.

Sevelsted, A., & Toubøl, J. (Eds.). (2022). *The power of morality in movements: Civic engagement in climate justice, human rights, and democracy*. Springer.

- Taylor, C. (2007). *A secular age*. Belknap Press of Harvard University Press.
- Vacant, A. (1908). *Dictionnaire de théologie catholique*. Letouzey et Ané.
- Witte, J., Jr. (1993). *Christianity and democracy in global context*. Westview Press

# La crítica rapsódica de Rawls al capitalismo: un fundamento antropológico para la democracia de propietarios

*Rawls's Rhapsodic Critique of Capitalism: An Anthropological Foundation for a Property-Owning Democracy*

**Javier Enrique Castillo Vallez<sup>67</sup>**

---

<sup>67</sup> Universidad de Chile. <https://orcid.org/0000-0002-4893-5425>. El financiamiento de este proyecto fue a través de la Agencia Nacional de Investigación y Desarrollo (ANID-PFCHA/Doctorado Nacional/Año 2022—21221670).

## Resumen

Este artículo sostiene que la crítica de John Rawls al capitalismo de bienestar, aunque aparentemente fragmentaria, revela un fundamento antropológico coherente basado en su constructivismo kantiano, operacionalizado metodológicamente a través del equilibrio reflexivo. Mediante un análisis teórico, se demuestra que la tensión entre autonomía individual y condiciones materiales se resuelve en la democracia de propietarios, que emerge como la realización institucional más fiel de la justicia como equidad. La investigación integra las perspectivas de O'Neill, Thomas, Malleson y Ron, y responde a las críticas de Kymlicka y Cohen desde el marco del equilibrio reflexivo amplio. La conclusión principal es que la democracia de propietarios representa la consecuencia lógica del liberalismo político rawlsiano, superando las limitaciones del capitalismo de bienestar y el socialismo liberal estatista, al institucionalizar las condiciones materiales necesarias para el desarrollo y ejercicio de los dos poderes morales de la persona.

**Palabras clave:** Democracia de propietarios; constructivismo kantiano; antropología política; predistribución; autonomía; equilibrio reflexivo.



## **Abstract**

This article argues that John Rawls's critique of welfare capitalism, while seemingly fragmented, reveals a coherent anthropological foundation based on his Kantian constructivism, operationalized through reflective equilibrium. Using theoretical analysis, it demonstrates that the tension between individual autonomy and material conditions is resolved in the property-owning democracy, which emerges as the most faithful institutional realization of justice as fairness. The research integrates perspectives from several scholars and responds to key criticisms from within the framework of wide reflective equilibrium. The main conclusion is that a property-owning democracy represents the logical consequence of Rawlsian political liberalism, as it institutionalizes the material conditions necessary for developing the person's two moral powers.

**Keywords:** Property-owning democracy; Kantian constructivism; political anthropology; predistribution; autonomy; reflective equilibrium.

## **Introducción**

John Rawls, en sus obras posteriores, desarrolla una crítica sustancial al capitalismo de bienestar por violar sistemáticamente el valor equitativo de las libertades políticas, permitir desigualdades extremas en la propiedad y

no reconocer un principio de reciprocidad para regular las desigualdades económicas (Rawls, 2001). Sin embargo, esta crítica podría ser calificada de “rapsódica” por carecer de un desarrollo sistemático y un hilo conductor explícito, presentándose más como observaciones dispersas que como una refutación integral.

Este artículo sostiene que, tras dicha apariencia fragmentaria, subyace un fundamento antropológico coherente enraizado en el constructivismo kantiano de Rawls (1980) y en el método del equilibrio reflexivo. A través de un análisis de esta fundamentación, se argumenta que la ambigüedad en la distinción entre lo formal y lo sustantivo no es una debilidad, sino la clave que permite interpretar tanto al socialismo liberal como a la democracia de propietarios como regímenes compatibles con la justicia como equidad. La hipótesis central postula que la crítica rawlsiana al capitalismo se articula a través de una tensión no resuelta entre autonomía individual y condiciones materiales, una tensión que encuentra su origen y resolución en la concepción de la persona como poseedora de dos poderes morales, concepción que es operacionalizada y justificada a través del equilibrio reflexivo. Esta tensión encuentra su resolución institucional en la “democracia de propietarios”, un régimen que realiza los requisitos de la autonomía plena al proporcionar la base material sin la cual la libertad se convierte en una mera formalidad, satisfaciendo así los

intereses de orden superior derivados de la estructura antropológica rawlsiana.

La pregunta de investigación que guía este trabajo es: ¿cómo la crítica fragmentaria de Rawls al capitalismo de bienestar revela un fundamento antropológico, articulado a través del constructivismo kantiano y el equilibrio reflexivo, que justifica la democracia de propietarios como régimen compatible con la justicia como equidad?

Para demostrar esta tesis, el artículo integra las perspectivas de Martin O'Neill (2012), Alan Thomas (2012, 2017), Tom Malleson (2014) y Amit Ron (2008), demostrando que la democracia de propietarios emerge como la realización institucional más coherente de la concepción rawlsiana de la persona moral. Asimismo, se incorporan las críticas de Kymlicka (2002) y Cohen (2008), junto con desarrollos del constructivismo kantiano en Rawls (1980), O'Neill (2015), Guyer (2024), Kaufman (2012), Korsgaard (2008) y Williams (2024), y una comprensión metodológica del equilibrio reflexivo (Knight, 2025), para evidenciar que la fundamentación antropológica resuelve las aparentes aporías de su teoría, cerrando la brecha entre lo correcto y lo bueno mediante una reestructuración predistributiva de la propiedad. Se sitúa explícitamente a Rawls dentro del debate liberal más amplio (Courtland et al., 2022). Esta contextualización refuerza la tesis central al mostrar que la democracia de propietarios no es una desviación del

liberalismo rawlsiano, sino su consecuencia lógica ante las insuficiencias del liberalismo clásico y los riesgos perfeccionistas del comprehensivo. Frente al liberalismo clásico, Rawls representa el “nuevo liberalismo” o liberalismo igualitario, el cual cuestiona que el mercado libre baste para garantizar una sociedad estable y justa (Courtland et al., 2022, sec. 2.2). A su vez, frente al liberalismo comprehensivo de Raz, el “giro político” de Rawls hacia un liberalismo no metafísico (Rawls, 1993) le permite defender la democracia de propietarios como una realización de la autonomía compatible con el pluralismo razonable, un pluralismo que se revela como consecuencia de la formalidad de la concepción antropológica misma (Fuentes & Castillo, 2024). Esta articulación permite, a su vez, reinterpretar la crítica comunitarista de Sandel (1982) para defender que la concepción de la persona en Rawls no es “desencarnada”, sino que especifica las condiciones formales y materiales necesarias para la agencia moral en una sociedad pluralista, condiciones cuya justificación se construye a través del equilibrio reflexivo. En definitiva, esta investigación concluye que justamente la democracia de propietarios es la encarnación institucional de un liberalismo político que toma en serio tanto la libertad como la igualdad, fundado en una antropología política robusta y metodológicamente coherente.

## **La crítica rapsódica de Rawls al capitalismo: fragmentos de un diagnóstico estructural**

La aparente desconexión de las críticas rawlsianas al capitalismo de bienestar oculta una profunda coherencia estructural que deriva directamente de su fundamento antropológico. Como demuestra O'Neill (2012), Rawls identifica tres déficits fundamentales en el Estado de Bienestar Capitalista (WSC<sup>68</sup>): primero, la violación sistemática del valor equitativo de las libertades políticas, donde la concentración oligárquica del capital distorsiona irreversiblemente el proceso democrático; segundo, el incumplimiento estructural de la igualdad equitativa de oportunidades, al institucionalizar la transmisión intergeneracional de ventajas mediante mecanismos hereditarios; y tercero, la sustitución del principio de reciprocidad por una mera lógica caritativa que niega la dignidad de los receptores (O'Neill, 2012, pp. 77-78). Estos déficits no son meras fallas empíricas, sino síntomas de una incompatibilidad profunda con la concepción rawlsiana de la persona como agente autónomo cuyos poderes morales requieren condiciones materiales para su ejercicio.

Esta tríada de críticas, aunque presentada de manera aparentemente desconectada, obedece a lo que Thomas

---

<sup>68</sup> Esta es la sigla para: Welfare State Capitalism. En adelante, usaré WSC para ahorrar espacio.

(2017) identifica como una ambigüedad fundacional en la distinción rawlsiana entre lo formal y lo sustantivo, una ambigüedad que la antropología del constructivismo kantiano resuelve. Rawls nunca abandonó completamente una “concepción formal” de la persona como mero titular de derechos, pero simultáneamente desarrolló una “concepción sustantiva” que exige condiciones materiales específicas para el ejercicio efectivo de la autonomía, condiciones que son precisadas a través del equilibrio reflexivo. Esta tensión conceptual solo encuentra resolución institucional en sus obras tardías mediante la adopción de la democracia de propietarios como marco económico-político preferido, por ser el que mejor satisface los intereses de orden superior de la persona moral. Como señala Thomas, Rawls desarrolló un “especificacionismo institucional” que trasciende el liberalismo político tradicional al incorporar preocupaciones republicanas sobre la dominación estructural (Thomas, 2017, p. 17). Este giro responde a una tensión interna en su proyecto: cómo garantizar que la igualdad política formal no sea socavada por desigualdades económicas sustantivas que generan relaciones de dominación, una tensión que el método del equilibrio reflexivo busca resolver ajustando principios, juicios y teorías de fondo.

Ahora bien, la supuesta “rapsodia” crítica es, en realidad, la expresión sintomática de una intuición antropológica profunda que busca su adecuada realización institucional a través del proceso de especificación que el

equilibrio reflexivo permite. Como señala Thomas (2017), “Rawls llegó a ver que sus puntos de vista solo podrían ser defendibles en su forma completamente especificada en un esquema de socialismo liberal de mercado o en una democracia de propietarios” (p. 19). La crítica al capitalismo de bienestar no es, por tanto, una colección de observaciones dispersas, sino los fragmentos de un diagnóstico estructural coherente: el WSC opera sobre una base de concentración inicial de capital que le impide garantizar un trasfondo estable de justicia a largo plazo, condenándolo a la inestabilidad inherente que Rawls temía, precisamente porque socava las bases sociales del autorrespeto y la autonomía efectiva. Thomas argumenta que la exclusión del estado de bienestar capitalista es la conclusión lógica del principio de reciprocidad rawlsiano, principio que a su vez se deriva de la estructura de la agencia moral (racional/razonable). El estado de bienestar “deja en su lugar estructuras de dominación e intenta paliar sus efectos”, una estrategia “inconsistente” con la preocupación rawlsiana por la no-dominación (Thomas, 2017, p. 20). Al aceptar una división fundamental de clases, perpetúa una relación de dependencia donde la redistribución es reactiva, políticamente contingente y vulnerable a la erosión por la dinámica de concentración del capital, imposibilitando así el desarrollo pleno de los dos poderes morales para una parte significativa de la ciudadanía.

Este diagnóstico estructural se clarifica al contrastarlo con la posición del liberalismo clásico desde la perspectiva de

la antropología rawlsiana. Para pensadores como Nozick (1974), cualquier patrón distributivo, incluido el igualitario, es violado constantemente por las transacciones libres de los individuos, y su mantenimiento requiere una interferencia estatal continua e inaceptable (Courtland et al., 2022, sec. 2.3). La réplica rawlsiana, implícita en su crítica al WSC y fundamentada en su constructivismo, es que el propio mercado “libre” del capitalismo de bienestar está estructuralmente sesgado por distribuciones iniciales arbitrarias de capital y dotaciones naturales que violan la igualdad equitativa de oportunidades y el principio de reciprocidad. Por lo tanto, la intervención estatal para establecer una “predistribución”<sup>69</sup> justa a través de una

---

<sup>69</sup> Atendiendo a una observación que me hizo un colega, Javier Fuentes, quiero hacer una aclaración en torno al concepto de “predistribución”. Como señala Fuentes (en comunicación personal), la predistribución podría entenderse como el acto de expropiación por parte del Estado, pero creo que en este contexto no debe confundirse con esta forma intensificada de redistribución estatal. Como se verá, mientras la redistribución corrige *ex post* desigualdades generadas por un marco institucional dado, la predistribución establece *ex ante* las reglas de fondo que determinan cómo se generan y distribuyen inicialmente los activos productivos. En la democracia de propietarios, mecanismos como la dotación universal de capital, los fondos de capital social, la regulación estructural del crédito y la tributación patrimonial no constituyen una expropiación de los propietarios existentes, sino una reconfiguración de las reglas de adquisición y transmisión de la propiedad (Thomas, 2017, pp. 216-234). Asimismo, como se verá, desde la antropología rawlsiana, la propiedad privada de los medios de producción no es un derecho natural prepolítico, sino una libertad institucionalmente configurada que solo se justifica si garantiza las condiciones materiales para el ejercicio de los dos poderes morales (Rawls, 2001; Courtland et al., 2022, sec. 2.3). La



democracia de propietarios no es una distorsión de la libertad, sino un requisito previo para un trasfondo verdaderamente justo sobre el cual puedan tener lugar transacciones libres y equitativas, un trasfondo que es condición de posibilidad para la autonomía efectiva. La crítica de Rawls al WSC es, en esencia, la constatación de que el liberalismo clásico, al ignorar la justicia de la estructura básica, consagra un orden que sistemáticamente viola sus propios ideales de libertad e igualdad de oportunidades, porque parte de una antropología insuficiente que no reconoce la dimensión material de los poderes morales.

### **El constructivismo kantiano como fundamento antropológico**

El constructivismo kantiano en Rawls representa mucho más que una metodología para derivar principios de justicia; constituye la expresión de una antropología política sustantiva y bien articulada. Como el propio Rawls (1980) explica, este enfoque "especifica una concepción particular de la persona como un elemento en un procedimiento razonable de construcción, cuyo resultado determina el contenido de los primeros principios de justicia" (p. 516). Esta afirmación revela la estructura profunda del proyecto rawlsiano: no se trata meramente de encontrar principios justos, sino de

---

predistribución es así la condición de legitimidad del sistema de propiedad privada, no una violación de derechos adquiridos.

derivarlos de una concepción antropológica específica. La persona moral, caracterizada por los dos poderes morales—la capacidad para un sentido de la justicia y la capacidad para formar, revisar y perseguir racionalmente una concepción del bien—junto con el ideal de autonomía plena, funciona como el núcleo normativo generativo de todo el sistema rawlsiano (Rawls, 1980). Leída a través de las herramientas analíticas de los "conceptos tópicos", esta concepción revela una estructura antropológica profunda que disuelve falsas dicotomías y exige una realización institucional específica (Fuentes & Castillo, 2024)<sup>70</sup>. La "posición original", desde esta perspectiva, es el procedimiento de construcción que media entre esta concepción antropológica y los principios de justicia, encarnando un "punto de vista social construido" (Korsgaard, 2008, p. 312). Este constructivismo se operacionaliza metodológicamente a través del equilibrio reflexivo (Knight, 2025).

### *Los poderes morales como estructura antropológica fundamental*

La concepción rawlsiana de la persona se articula alrededor de dos poderes morales que son estructuras antropológicas normativas: la capacidad para un sentido de la justicia y la capacidad para formar, revisar y perseguir racionalmente una concepción del bien (Rawls, 1980, p. 525).

---

<sup>70</sup> Particularmente los pares forma/materia e interno/externo.

Estos poderes definen lo que significa ser un agente moral en un contexto social. Encuentran su expresión epistemológica primaria en los *juicios considerados*, que constituyen el punto de partida del equilibrio reflexivo<sup>71</sup>.

Esta conexión es crucial: los juicios considerados no son opiniones arbitrarias, sino la manifestación concreta de nuestras capacidades morales en circunstancias ideales de deliberación. Estos poderes no son atributos contingentes, sino constituyentes de la personalidad moral en una sociedad democrática. Rawls conecta esta concepción directamente con la tradición kantiana al señalar que "la concepción de la persona, el procedimiento y los primeros principios deben relacionarse de cierta manera" (1980, p. 516). Lo crucial aquí es que estos poderes no son meramente formales: generan intereses de orden superior que requieren condiciones materiales específicas para su realización. El interés de orden superior en el ejercicio del sentido de la justicia exige instituciones que hagan posible comprender y aplicar principios justos; el interés en desarrollar y ejercer la capacidad para una concepción del bien exige condiciones que

---

<sup>71</sup> Para ser considerados, los juicios deben cumplir tres condiciones: (1) capacidad para tomar la decisión correcta (estar razonablemente informado), (2) estar en circunstancias que lo permitan (sin miedo o alteración), y (3) estar motivado para hacerlo (sin intereses creados) (Rawls, 1971, p. 47, citado en Knight, 2025). Son "aquellos juicios en los que nuestras capacidades morales tienen más probabilidad de mostrarse sin distorsión".

permitan la formación, revisión y persecución racional de planes de vida.

Esta antropología tiene implicaciones radicales para la estructura social. Como señala Rawls, estos intereses de orden superior son "supremamente regulativos así como efectivos. Esto implica que, siempre que las circunstancias sean relevantes para su cumplimiento, estos intereses gobiernan la deliberación y la conducta" (1980, p. 525). La sociedad justa, por tanto, no es aquella que meramente no interfiere con estos poderes, sino aquella que positivamente crea las condiciones para su desarrollo y ejercicio. Esto marca una diferencia crucial con concepciones liberales más minimalistas: la antropología rawlsiana exige un marco institucional robusto que active y sostenga estas capacidades constitutivas de la personalidad moral.

### *La distinción racional/razonable como estructura de la agencia moral*

La crucial distinción rawlsiana entre lo racional (la búsqueda informada de una concepción del bien) y lo razonable (la disposición a aceptar y proponer términos justos de cooperación) constituye los pilares de una antropología política que demanda condiciones materiales específicas para su realización efectiva. Rawls es explícito sobre esta distinción: "El Razonable presupone y subordina a lo Racional. Define los términos justos de cooperación

aceptables para todos dentro de algún grupo de personas identificables por separado, cada una de las cuales posee y puede ejercer los dos poderes morales" (1980, p. 530). Esta distinción no es meramente metodológica sino antropológicamente constitutiva: define la estructura misma de la agencia moral en contexto social.

Lo racional corresponde al ejercicio del segundo poder moral: la capacidad para formar y perseguir una concepción del bien. Lo razonable corresponde al ejercicio del primer poder moral: la capacidad para un sentido de la justicia. La subordinación de lo racional a lo razonable —"en una doctrina kantiana limita absolutamente los fines finales que pueden perseguirse" (p. 530)— refleja una antropología donde la autonomía colectiva tiene prioridad sobre la mera racionalidad instrumental individual. Esta estructura explica por qué Rawls insiste en que:

la unidad de la razón práctica se expresa definiendo lo Razonable para enmarcar lo Racional y subordinarlo absolutamente; es decir, los principios de justicia acordados son léxicamente prioritarios en su aplicación en una sociedad bien ordenada a los reclamos del bien. (p. 532)

Ahora bien, el equilibrio reflexivo amplio (*wide reflective equilibrium*) es el proceso metodológico mediante

el cual esta distinción estructural se traduce en principios de justicia concretos<sup>72</sup>.

Como argumenta Thomas (2012), sin una base material de independencia económica, esta estructura antropológica se vuelve disfuncional: la autonomía se convierte en una mera formalidad, una promesa vacía de libertad. Esta intuición es profundamente kantiana en su inspiración, pero radicalmente social en su ejecución.

### *El constructivismo como alternativa a fundamentaciones epistemológica*

Las interpretaciones contemporáneas del constructivismo rawlsiano, aunque divergen en sus énfasis, coinciden en subrayar su potencial crítico y antropológico. Por ejemplo, Guyer (2024) destaca cómo Rawls se apropia de la autonomía kantiana para trasplantarla a un contexto social, mientras que O'Neill (2015) argumenta que Rawls abandona

---

<sup>72</sup> Como describe Knight (2025), el equilibrio reflexivo amplio contrasta con el equilibrio reflexivo estrecho (narrow reflective equilibrium) en varios aspectos fundamentales: (1) considera un amplio rango de principios, no solo aquellos que se asemejan a nuestros juicios iniciales; (2) permite la revisión de nuestros juicios iniciales; (3) considera argumentos a favor y en contra de los principios; (4) incorpora "dispositivos de representación" como la posición original; y (5) integra teorías de fondo morales y no morales. Este proceso de ajuste mutuo entre juicios y principios bajo el escrutinio de argumentos y teorías es lo que permite que la estructura antropológica de lo racional y lo razonable se concrete en principios de justicia específicos.

el universalismo kantiano en favor de un constructivismo político restringido. Ambas lecturas, no obstante, reconocen que el kantismo de Rawls proporciona un marco potente para criticar las estructuras sociales que dificultan la autonomía. El texto de 1980 es crucial a este respecto, pues muestra a Rawls explicitando las raíces kantianas de su concepción y distanciándose de las doctrinas que postulan un orden moral independiente.

En él, Rawls contrasta explícitamente su constructivismo con el intuicionismo racional, que busca "verdades morales interpretadas como fijadas por un orden previo e independiente de objetos y relaciones, ya sea natural o divino" (p. 519). Frente a esto, "el constructivismo kantiano sostiene que la objetividad moral debe entenderse en términos de un punto de vista social construido adecuadamente que todos puedan aceptar. Aparte del procedimiento de construcción de los principios de justicia, no hay hechos morales" (p. 519). Esta afirmación radical tiene profundas implicaciones antropológicas: la moralidad no es un descubrimiento, sino una construcción que procede de una concepción específica de la persona.

Korsgaard (2008) profundiza esta crítica al realismo moral, exponiendo las limitaciones internas de su "modelo de aplicación" del conocimiento normativo. El equilibrio reflexivo implementa esta visión constructivista como una metodología coherentista, donde la justificación emerge de la

coherencia del sistema total de creencias del deliberador (Knight, 2025). Esta autora señala que el constructivismo kantiano redefine la objetividad moral no como correspondencia con un orden independiente, sino como el resultado de un punto de vista social construido que encarna la autonomía colectiva. Dicho punto de vista exige, como condición de su propia coherencia, que las instituciones sociales permitan a todos los ciudadanos desarrollar y ejercitar sus poderes morales, lo que a su vez requiere un acceso universal a los recursos productivos. La democracia de propietarios, al predistribuir el capital, institucionaliza este requisito: convierte el acceso a los recursos en una condición material de la autonomía colectiva, un bien primario en sentido rawlsiano fuerte, necesario para el desarrollo y ejercicio de los dos poderes morales (pp. 525-527).

El enfoque constructivista de Rawls, al generar principios a partir de una concepción de la persona y un procedimiento, se defiende así de otras posiciones metaéticas clásicas<sup>73</sup>. En particular, se enfrenta a las objeciones de G.A.

---

<sup>73</sup> La metaética se ocupa, como se indica, del estatus, fundamentos y alcance de los conceptos morales, operando en un nivel de abstracción que a menudo se denomina de “segundo orden”. Si bien John Rawls es conocido ante todo por su teoría normativa de la justicia como equidad, su obra aborda de manera sostenida cuestiones que podrían denominarse contemporáneamente “metaéticas”. Por ejemplo, en sus conferencias sobre historia de la filosofía moral y en escritos como “Kantian Constructivism in Moral Theory” (1980), aquí analizado, Rawls propone una posición constructivista que busca articular cómo pueden justificarse principios objetivos sin presuponer un realismo moral de corte



Cohen, quien atacó los fundamentos del constructivismo político. Cohen argumentó que el constructivismo rawlsiano, al permitir que "factores ajenos" a la justicia pura (como consideraciones de estabilidad o factibilidad) influyan en la derivación de principios, confunde los fundamentos de la justicia con reglas de regulación meramente instrumentales (Cohen, 2008; Williams, 2024, sec. 7.2). Para él, los principios de justicia fundamentales deben ser *fact-independent*; es decir, no depender de contingencias empíricas.

La réplica implícita en la antropología rawlsiana, operacionalizada mediante el equilibrio reflexivo, es que esta distinción rígida es insostenible en la práctica política. La concepción de la persona como poseedora de los dos poderes morales no es un hecho bruto, sino una idea normativa arraigada en la cultura pública de una sociedad democrática, cuya justificación se construye en un proceso de equilibrio reflexivo que sí incorpora teorías empíricas y consideraciones de estabilidad (Rawls, 1980; Knight, 2025). La "objección de los factores ajenos" pierde fuerza cuando se reconoce que la justicia, en sentido político rawlsiano, tiene como objeto la estructura básica de la sociedad y aspira a ser el foco de un consenso estable entre doctrinas comprehensivas razonables.

---

intuicionista. Dicha propuesta constituye una intervención directa en debates sobre la naturaleza de la verdad ética, el estatus de los principios y la relación entre la corrección moral y la racionalidad práctica; todos asuntos que actualmente se sitúan bajo dicha denominación.

Los requisitos de publicidad y estabilidad no son "ajenos", sino constitutivos de lo que significa que un principio de justicia sea practicable y justificable para personas libres e iguales en condiciones de pluralismo razonable (Rawls, 1993). Así, el constructivismo rawlsiano redefine la naturaleza y el ámbito de aplicación de los principios de justicia de manera política, no metafísica, una respuesta cuya metodología es el equilibrio reflexivo amplio.

### *La autonomía como principio unificador*

Siguiendo con el análisis del artículo de Rawls de 1980, se puede apreciar que la perspectiva kantiana sobre la autonomía, reinterpretada por Rawls, permite superar la dicotomía entre liberalismo político (neutralista) y perfeccionista. La clave es aplicar la distinción forma/materia como puntos de vista correlativos (Fuentes & Castillo, 2024). Según esto, la capacidad para una concepción del bien provee la materia de una vida; la capacidad para un sentido de la justicia provee la forma de reciprocidad bajo la cual esa materia se organiza (Rawls, 1980). Korsgaard (2008) refuerza esta visión al comparar los conceptos normativos con conceptos de artefactos, que nombran soluciones a problemas prácticos más que describen realidades independientes. El equilibrio reflexivo es el proceso que articula esta distinción, evaluando dispositivos como la posición original por su capacidad para generar principios coherentes con nuestros juicios y teorías (Knight, 2025). La perspectiva kantiana sobre

la autonomía, tal como es adoptada y reinterpretada por Rawls, permite resolver la aparente disyuntiva entre liberalismo político (neutralista) y perfeccionista que ha dominado la filosofía política contemporánea (Fuentes & Castillo, 2024). El liberalismo político (Rawls, 1993) pretende ser neutral respecto de las concepciones sustantivas del bien, defendiendo una autonomía meramente formal como parte de un consenso superpuesto. El liberalismo perfeccionista (Raz, 1986), en cambio, sostiene que el Estado debe promover activamente la autonomía como un valor constitutivo de una vida valiosa, lo que lo acerca a una doctrina comprehensiva particular.

Por tanto, el Estado es "perfeccionista" respecto a la forma de la agencia (la autonomía como *cómo*), pero "neutral" respecto a la materia de los proyectos de vida (Fuentes & Castillo, 2024). La clave para disolverla radica en aplicar la distinción forma/materia como puntos de vista correlativos y no como entidades separadas (Fuentes & Castillo, 2024; Wieland, 1992). Rawls proporciona la base para esto en su descripción de los poderes morales. Así, la capacidad para una concepción del bien (lo racional) provee la materia —los fines, proyectos, afectos— de una vida humana. Por su parte, la capacidad para un sentido de la justicia (lo razonable) provee la forma —la estructura de reciprocidad y universalidad— bajo la cual esa materia debe

organizarse para ser compatible con una sociedad de iguales<sup>74</sup>.

### *Pluralismo como consecuencia de la formalidad autónoma*

Rawls, en su "giro político", temió que una fundamentación kantiana comprensiva de la autonomía vulnerara el pluralismo razonable. Sin embargo, como argumentan Fuentes y Castillo (2024), el pluralismo no es

---

<sup>74</sup> El Estado no necesita adherir a una concepción sustantiva particular del bien (una materia específica) para promover la autonomía. Su tarea es garantizar las condiciones formales y materiales que hacen posible que toda concepción del bien razonable pueda ser perseguida autónomamente, es decir, de un modo propio. Esto implica que la neutralidad del Estado no es sobre la autonomía misma (que es el principio formal de la agencia), sino sobre los contenidos particulares (materia) que los individuos eligen para sus vidas. El Estado es, por tanto, "perfeccionista" respecto de la forma de la agencia (la autonomía como cómo), pero "neutral" respecto de la materia específica de los proyectos de vida (las concepciones del bien como qué). Como señalan Fuentes y Castillo (2024), un liberalismo así entendido puede llamarse "formal" o incluso "trascendental", "ya que trata con las condiciones de posibilidad de la autonomía desde un punto de vista político".

Esta superación de la dicotomía es posible porque, como muestran Castillo y Fuentes (2024), los conceptos de forma y materia son iterativos: lo que desde una perspectiva es forma, desde otra puede ser materia. Rawls ilustra esta iteratividad en su discusión sobre los bienes primarios. Estos son la materia (medios de propósito general) necesaria para perseguir cualquier concepción del bien, pero a su vez, su distribución justa es la forma que debe tomar la estructura básica para realizar la justicia como equidad. "Los bienes primarios se seleccionan preguntando qué cosas son generalmente necesarias como condiciones sociales y medios de propósito general para permitir a los seres humanos realizar y ejercer sus poderes morales y perseguir sus fines finales" (p. 526).

solo un "hecho" social que Rawls debe aceptar (Rawls, 1993), sino una consecuencia necesaria de la formalidad de la autonomía. Si la ley moral (o los principios de justicia) no prescribe un fin sustantivo (materia), sino la forma de la legislación universal (el cómo), entonces el espacio para una pluralidad de fines legítimos (materias) se abre necesariamente. Si los principios prescriben la forma de la legislación (el cómo), y no un fin sustantivo (la *materia*), se abre necesariamente un espacio para una pluralidad de fines legítimos (Fuentes & Castillo, 2024).

El equilibrio reflexivo proporciona la metodología para manejar este pluralismo en la justificación moral. Como señala Knight (2025), una objeción común al equilibrio reflexivo es que diferentes personas pueden alcanzar equilibrios diferentes, lo que podría sugerir relativismo. Sin embargo, defensores como Scanlon argumentan que "aceptar que alguna persona está bien informada y ha llevado a cabo la búsqueda del equilibrio reflexivo concienzudamente no compromete a uno siquiera con la conclusión de que esta persona está justificada en aceptar los principios que resultan" (Scanlon 2003, p. 152, citado en Knight 2025). La existencia de equilibrios divergentes no implica relativismo, sino que invita a investigar las razones de las divergencias: ¿diferentes juicios iniciales? ¿diferentes ajustes? ¿consideración de diferentes principios? Este proceso dialéctico es parte integral del equilibrio reflexivo como método de justificación en contextos pluralistas.

### *La posición original como dispositivo de representación*

La posición original es el procedimiento de construcción que media entre la concepción de la persona y los principios de justicia (Rawls, 1980, p. 520). Desde el equilibrio reflexivo, es un "dispositivo de representación" cuya utilidad se juzga por su ajuste a nuestros juicios (Knight, 2025). Su "velo de ignorancia denso" representa la independencia de las personas respecto a sus fines contingentes (Rawls, 1980, pp. 548-550). Esta independencia es un aspecto clave de la libertad: las personas se conciben como capaces de revisar sus fines (Rawls, 1980, p. 522). Si las partes en la posición original buscan asegurar condiciones para cualquier concepción del bien, necesitarán garantizar recursos materiales que hagan posible esa revisión autónoma, conectando con la necesidad de una distribución amplia de la propiedad (Rawls, 1980, p. 526).

### *Los tres niveles de publicidad y la objetividad construida a través del equilibrio reflexivo*

La publicidad en Rawls tiene tres niveles que refuerzan la fundamentación antropológica. En el primer nivel, "significa que la sociedad está efectivamente regulada por principios públicos de justicia; es decir, todos aceptan y saben que los demás aceptan los mismos principios, y este conocimiento a su vez es públicamente reconocido" (1980, p. 537). El segundo nivel "se refiere a las creencias generales a la

luz de las cuales los primeros principios de justicia mismos pueden aceptarse, es decir, la teoría de la naturaleza humana y de las instituciones sociales en general" (1980, p. 537). El tercer nivel "tiene que ver con la justificación completa de la concepción pública de justicia como se presentaría en sus propios términos" (1980, p. 537).

Esta exigencia de publicidad se conecta directamente con el equilibrio reflexivo público. Como explica Knight (2025), Rawls evoluciona desde un equilibrio reflexivo individualista en *Teoría de la Justicia* hacia un "consenso superpuesto" en *Liberalismo Político*. En este marco, "los equilibrios reflexivos de adherentes de cada una de estas doctrinas comprensivas pueden converger en un conjunto específico de principios liberales, cada uno por diferentes razones" (Knight 2025). Rawls se refiere a esto como equilibrio reflexivo completo (full reflective equilibrium): "la misma concepción es afirmada en los juicios considerados de todos" (Rawls 2005: 384 n. 16, citado en Knight 2025)<sup>75</sup>.

---

<sup>75</sup> Rawls concluye que "la objetividad no está dada por 'el punto de vista del universo', para usar la frase de Sidgwick. La objetividad debe entenderse por referencia a un punto de vista social adecuadamente construido" (p. 570). La democracia de propietarios gana su justificación objetiva no por corresponder a una verdad moral independiente, sino por ser la realización institucional del punto de vista que personas libres e iguales construirían para sí mismas desde la posición original, justificada a través del equilibrio reflexivo público y completo.

## *El equilibrio reflexivo como realización metodológica de la antropología rawlsiana*

El equilibrio reflexivo, particularmente en su forma amplia y pública, constituye la realización metodológica concreta de la antropología rawlsiana. Como resume Knight (2025), este método implica "el ajuste mutuo de principios y juicios a la luz de argumentos y teorías relevantes". Este proceso dialéctico opera en múltiples dimensiones que corresponden a los diferentes niveles de la antropología rawlsiana:

- En el nivel constitutivo, los juicios considerados expresan operativamente los dos poderes morales. Como señala Knight, estos juicios son "aquellos en los que nuestras capacidades morales tienen más probabilidad de mostrarse sin distorsión". El requisito de que sean hechos en circunstancias ideales de deliberación (con habilidad, oportunidad y deseo de juzgar correctamente) garantiza que reflejen fielmente nuestros poderes morales básicos.
- En el nivel estructural, el proceso de ajuste mutuo entre juicios particulares y principios generales realiza la relación entre lo racional y lo razonable. Los principios propuestos deben no solo capturar nuestros juicios existentes, sino también extenderse a nuevos casos y sobrevivir al escrutinio de argumentos contrapuestos. Este proceso asegura que lo razonable (la forma de



reciprocidad) discipline a lo racional (la persecución de concepciones del bien).

- En el nivel procedimental, la posición original y otros "dispositivos de representación" funcionan como herramientas heurísticas dentro del equilibrio reflexivo. Su validez no es independiente, sino que se juzga por su capacidad para generar principios que, a su vez, se ajusten coherentemente a nuestro sistema de juicios y teorías. Como señala Knight, "si encontramos útil cualquier dispositivo, es una cuestión de si se ajusta a nuestros juicios en equilibrio reflexivo amplio".
- En el nivel epistémico, el carácter coherentista del equilibrio reflexivo implementa la visión constructivista de la objetividad moral. En lugar de buscar correspondencia con hechos morales independientes, la justificación emerge de la coherencia del sistema total de creencias del deliberador, incluyendo juicios morales, principios, argumentos filosóficos y teorías de fondo tanto morales como no morales.
- En el nivel institucional, la conexión entre el equilibrio reflexivo y las exigencias de publicidad asegura que los principios resultantes puedan servir como base para una justificación pública en una sociedad pluralista. El equilibrio reflexivo completo (full reflective equilibrium) busca precisamente la convergencia entre diferentes

doctrinas comprensivas hacia principios políticos compartidos.

Este método responde a objeciones como las de relativismo o indeterminación, ofreciendo un "campo de juego nivelado" para la justificación (Knight, 2025)<sup>76</sup>.

---

<sup>76</sup> Contra la objeción de que los juicios iniciales carecen de credibilidad, el equilibrio reflexivo amplio somete estos juicios a un riguroso escrutinio teórico. Contra la objeción de relativismo, la posibilidad de equilibrios divergentes no invalida el método, sino que invita a la investigación de las razones de las divergencias. Contra la objeción de indeterminación, el hecho de que puedan existir múltiples equilibrios no es un defecto, sino un reconocimiento de la complejidad moral. Contra la objeción de sesgos, el equilibrio reflexivo permite incorporar teorías psicológicas sobre sesgos cognitivos como parte de las teorías de fondo relevantes.

Como concluye Knight citando a Scanlon, el equilibrio reflexivo puede parecer "cercano a vacío" porque permite "incorporar lo que podrían haber sido vistos como métodos alternativos de justificación dentro de él". Pero esto es una característica, no un error: "Lo que el método del equilibrio reflexivo prescribe es, por así decirlo, un campo de juego nivelado de justificación intuitiva en el que principios y juicios de todos los niveles de generalidad deben competir por nuestra lealtad". Esta apertura metódica es precisamente lo que permite al equilibrio reflexivo realizar la riqueza de la antropología rawlsiana sin reducirla a dogmatismos apriorísticos.

## **Conclusión: la antropología rawlsiana como fundamento de la democracia de propietarios a través del equilibrio reflexivo**

El constructivismo kantiano en Rawls, operacionalizado a través del método del equilibrio reflexivo, proporciona un fundamento antropológico robusto para la democracia de propietarios. Esta fundamentación opera en múltiples niveles que se refuerzan mutuamente:

- **Nivel constitutivo:** los dos poderes morales, expresados en juicios considerados, definen lo que significa ser una persona en una sociedad democrática, generando intereses de orden superior que exigen condiciones materiales específicas.
- **Nivel estructural:** la distinción entre lo racional y lo razonable, elaborada mediante el ajuste mutuo entre juicios particulares y principios generales en el equilibrio reflexivo, organiza la agencia moral subordinando la racionalidad instrumental a la reciprocidad justa.
- **Nivel procedimental:** la posición original, como dispositivo de representación justificado dentro del equilibrio reflexivo amplio, media entre la concepción abstracta de la persona y los principios concretos de justicia, exigiendo condiciones que permitan la independencia y la autonomía.

- **Nivel institucional:** los bienes primarios, especialmente los recursos materiales, son necesarios para el ejercicio de los poderes morales, lo que requiere una distribución amplia de la propiedad justificada a través del equilibrio reflexivo con apoyo en teorías económicas y sociales de fondo.
- **Nivel epistémico:** la objetividad se redefine como una construcción social a través del equilibrio reflexivo coherentista, y no como el descubrimiento de verdades morales independientes.
- **Nivel público:** el equilibrio reflexivo completo busca la convergencia entre doctrinas comprehensivas diversas, asegurando que los principios resultantes puedan servir como base de justificación pública en una sociedad pluralista.

Esta antropología no cae en el perfeccionismo sustantivo, sino que especifica las condiciones de posibilidad de una sociedad justa entre personas libres e iguales, que es el núcleo del liberalismo político rawlsiano. La democracia de propietarios emerge así no como una opción política más, sino como la realización institucional necesaria de esta antropología política, justificada mediante el equilibrio reflexivo. Como concluye Rawls, encontrar una salida al impasse entre libertad e igualdad "define la tarea práctica inmediata de la filosofía política" (1980, p. 572). La

democracia de propietarios, fundada en esta antropología constructivista kantiana, ofrece precisamente esa salida, cerrando la brecha entre el ideal normativo y las condiciones de su realización.

La fundamentación antropológica a través del constructivismo y el equilibrio reflexivo permite, por tanto, responder a críticas internalistas y externalistas al proyecto rawlsiano. Frente al realismo moral o el intuicionismo racional, que buscan verdades morales en un orden independiente, el constructivismo muestra que la objetividad de la justicia se construye desde un punto de vista social compartido (Rawls, 1980; Korsgaard, 2008). Frente a la crítica de Cohen, muestra que la incorporación de consideraciones de estabilidad y publicidad no es una corrupción de la justicia, sino una exigencia de su realización práctica en un mundo de escasez moderada y pluralismo. El equilibrio reflexivo es el método que hace coherente esta incorporación, al permitir un ajuste mutuo entre principios abstractos, juicios considerados y teorías de fondo sobre la psicología moral y la organización social (Knight, 2025). En este proceso, la democracia de propietarios emerge no solo como la especificación institucional de los principios de justicia, sino también como la realización de un ideal de autonomía que puede ganarse el apoyo de una pluralidad de doctrinas comprehensivas. Esta construcción evita tanto el perfeccionismo comprehensivo como el neutralismo liberal minimalista, al tiempo que supera las limitaciones del

capitalismo de bienestar y del socialismo estatista. De esta manera, la fortaleza del enfoque rawlsiano reside en su capacidad para integrar exigencias normativas sustantivas (derivadas de su antropología) con restricciones metodológicas y prácticas (trabajadas a través del equilibrio reflexivo), ofreciendo una teoría política.

### *La ambigüedad en la propiedad y su resolución institucional*

La ambigüedad en el tratamiento de la propiedad de los medios de producción constituye quizás el punto más revelador de la antropología política rawlsiana y de su método de equilibrio reflexivo. Esta ambigüedad no es una debilidad teórica, sino la consecuencia de un enfoque constructivista que, partiendo de una concepción formal de la persona, debe especificar institucionalmente las condiciones materiales para sus poderes morales, un proceso que admite más de una vía coherente. Esta ambigüedad permite a Rawls considerar tanto el socialismo liberal (propiedad colectiva o social) como la democracia de propietarios (propiedad individual dispersa) como regímenes potencialmente justos (Malleon, 2014), siempre que pasen la prueba del equilibrio reflexivo amplio al satisfacer los principios derivados de la concepción de la persona.

Rawls (2001) es explícito al negar que se pueda deducir un derecho a la propiedad privada de los medios de producción de sus primeros principios. Su enfoque se declara

“puramente político”: busca un terreno de acuerdo en la cultura pública de una democracia, evitando comprometerse con fundamentos filosóficos metafísicos sobre la propiedad. Como señala Thomas (2017), citando a Rawls (1993):

[E]l objetivo de la justicia como equidad como concepción política es resolver el punto muerto en la tradición democrática en cuanto a la forma en que deben organizarse las instituciones sociales si van a ajustarse a la libertad e igualdad de los ciudadanos como personas morales. El argumento filosófico por sí solo es poco probable que convenza a cualquiera de los dos bandos de que el otro es correcto en una cuestión como la de la propiedad privada o social de los medios de producción. Parece más fructífero buscar las bases del acuerdo implícitas en la cultura pública de una sociedad democrática. (pp. 338–339, citado en Thomas, 2017, p. 39)

Sin embargo, esta neutralidad declarada es solo aparente cuando se la examina desde la antropología constructivista. Al exigir que el régimen económico garantice el valor equitativo de las libertades políticas, la igualdad equitativa de oportunidades y el principio de diferencia, Rawls establece criterios sustantivos que cualquier régimen de propiedad debe cumplir para ser coherente con la concepción de la persona. Lo que importa, desde la perspectiva del constructivismo kantiano y del equilibrio

reflexivo, no es la forma de propiedad per se, sino si esta garantiza las condiciones materiales para la autonomía y la no-dominación, es decir, para el ejercicio efectivo de los dos poderes morales.

En este contexto, la propiedad, en particular la propiedad productiva, no es un fin en sí mismo, sino la condición material necesaria para la realización efectiva de la autonomía. La mera libertad interna o la posesión formal de derechos son insuficientes si el individuo carece de los recursos materiales para trazar y seguir su propio plan de vida sin depender de la voluntad arbitraria de otros. La democracia de propietarios, al dispersar universalmente la propiedad del capital mediante mecanismos predistributivos, proporciona la base económica que debe ser configurada por la autonomía. Resuelve así la tensión entre lo formal y lo sustantivo que Thomas identifica: la autonomía interna (la capacidad para concebir un plan de vida) se ve posibilitada y potenciada por una base externa de independencia material. Esta resolución es la que emerge como más coherente en un equilibrio reflexivo que incorpora teorías sobre la estabilidad, el pluralismo y los peligros de la concentración del poder.

Esta perspectiva permite una respuesta robusta a la crítica de Kymlicka (2002) sobre la insensibilidad del principio de diferencia a las elecciones individuales. La democracia de propietarios, al dispersar la propiedad ex-ante mediante mecanismos predistributivos, mitiga este problema



al crear una distribución inicial justa de activos productivos. Esto permite que las desigualdades posteriores reflejen en mayor medida elecciones responsables, en lugar de ventajas heredadas o dotaciones naturales no compensadas (Thomas, 2017). La estructura institucional de la democracia de propietarios asegura así una mayor “sensibilidad a la ambición” mientras se mantiene rigurosamente “insensible a la dotación”, realizando de manera más precisa la concepción de la persona como responsable de sus fines pero no de sus dotaciones iniciales.

*La superioridad de la democracia de propietarios sobre el socialismo de mercado obligatorio desde la perspectiva antropológica*

Aunque Rawls presentó ambas opciones como equivalentes, Thomas, desde su síntesis liberal-republicana, argumenta a favor de la superioridad de la democracia de propietarios, especialmente frente al socialismo de mercado “obligatorio” o “mandatario”. Su crítica es severa y revela la profundidad del compromiso rawlsiano con el pluralismo y la libertad, valores que se derivan directamente de la formalidad de la concepción antropológica.

Thomas argumenta que las formas obligatorias de socialismo de mercado (donde la propiedad cooperativa es la única forma legal permitida) son inherentemente explotadoras y anti-liberales. La explotación, en términos

rawlsianos de falta de reciprocidad, ocurre cuando “el trabajador es compelido a ser parte de una relación que sistemáticamente fracasa en ser recíproca” (Thomas, 2017, p. 55, citando a Arnold, 1994, p. 58). Al eliminar la propiedad privada del capital y concentrarlo en manos del Estado como arrendador único, se crea un nuevo y masivo problema de agente-principal donde los trabajadores-cooperativistas tienen incentivos para explotar el capital arrendado, subinvertir y apropiarse de beneficios a expensas de las generaciones futuras (Thomas, 2017, pp. 229-234). Este sistema contradice el interés de orden superior en la justicia y la estabilidad.

Además, este sistema “amenaza la libertad de ocupación” al debilitar el mercado laboral y reducir la movilidad y el dinamismo empresarial (Thomas, 2017, p. 219). Así, forzar a todos a trabajar en empresas socializadas sin permitir alternativas viola la autonomía individual y el “hecho del pluralismo razonable”, imponiendo una concepción comprehensiva del bien laboral que es inadmisibles en un estado liberal (Thomas, 2017, pp. 262-264). Esto contradice la neutralidad del Estado respecto a la materia de las concepciones del bien, una neutralidad que es corolario de la formalidad de la autonomía rawlsiana.

Frente a esto, la democracia de propietarios es superior porque no impone una forma económica, sino que crea las condiciones materiales (la propiedad dispersa) para que

florezca un pluralismo de formas organizativas, respetando así la estructura racional/razonable de la agencia. La estrategia es “dar a cada ciudadano una libertad de mercado significativa primero, implementar los tres principios de justicia juntos, y podemos esperar razonablemente un floreciente rango de organizaciones económicas” (Thomas, 2017, p. 218). Esta diversidad refleja el valor republicano de dispersar el poder: “el poder disperso sostiene una gobernanza poliárquica y así protege la libertad individual” (Thomas, 2017, p. 216). La democracia de propietarios fomenta un “mancomunidad pluralista” (pluralistic commonwealth) con una rica sociedad civil de asociaciones voluntarias —sindicatos, cooperativas, fondos éticos— donde los ciudadanos ejercen su autonomía (Thomas, 2017, pp. 277-279). Este pluralismo institucional es la encarnación social del pluralismo razonable, y es coherente con una antropología que prioriza la forma de la cooperación justa sobre cualquier contenido económico particular.

La ambigüedad rawlsiana sobre la propiedad y su resolución a favor de una amplia dispersión de la misma marca una diferencia crucial con el liberalismo clásico y libertario. Para el liberalismo clásico (Hayek, Nozick) existe una conexión inseparable entre la libertad individual y un orden robusto de propiedad privada de los medios de producción. Para Rawls, sin embargo, la propiedad privada de los medios de producción no es un derecho básico. Es una libertad de los ciudadanos, pero está sujeta a los requisitos de

los dos principios de justicia (Courtland et al., 2022, sec. 2.3). La democracia de propietarios es, por tanto, la antítesis del libertarianismo de Nozick. Mientras Nozick ve cualquier redistribución (o predistribución) como una violación de los derechos de propiedad, Rawls argumenta que un sistema de propiedad privada solo es justificable si está estructurado para servir a los fines de la justicia, es decir, para garantizar las condiciones para el ejercicio de los poderes morales. La predistribución en una democracia de propietarios no es, entonces, una violación de la propiedad, sino la condición de su legitimidad desde el punto de vista de la antropología rawlsiana y el equilibrio reflexivo que la sustenta.

*La democracia de propietarios como realización de la reciprocidad y los intereses de orden superior*

La democracia de propietarios encarna el principio de reciprocidad de manera más completa que cualquier alternativa institucional, satisfaciendo simultáneamente las condiciones de la autonomía plena y la razonabilidad (Thomas, 2017; O'Neill, 2012). Al implementar una "predistribución" sistemática del capital, este régimen transforma cualitativamente la naturaleza de los incentivos en el mercado, respondiendo a la crítica de Cohen (2008) sin recurrir a un "ethos igualitario" suplementario, pues la institución misma materializa y asegura ese ethos. Esto satisface directamente los intereses de orden superior de la

persona de contar con instituciones estables que apoyen su sentido de la justicia.

Como argumenta Thomas (2012, 2017), en una democracia de propietarios con propiedad ampliamente dispersa, los incentivos que emergen son predominantemente "compensatorios" (por desutilidades reales o esfuerzos extraordinarios) y no "extorsivos" (producto de un poder de negociación estructuralmente desigual). La razón es que todos los ciudadanos poseen una base material independiente que les permite negociar en condiciones de mayor igualdad. La estructura institucional garantiza así la justicia de fondo, cerrando el "espacio libre de justicia" que Cohen identificaba en el marco rawlsiano tradicional, porque el diseño mismo de la estructura básica hace que las acciones individuales dentro de ella tiendan a producir resultados justos.

Esta interpretación institucional de la reciprocidad proporciona una respuesta robusta a la crítica persistente de G.A. Cohen (2008) respecto a la permisividad del principio de diferencia con los incentivos. Cohen argumentaba que el constructivismo rawlsiano, al limitarse a la estructura básica, dejaba un "espacio libre de justicia" en la esfera personal, permitiendo que los talentosos exigieran incentivos extras para producir más, incluso cuando esa demanda explota su posición ventajosa. La democracia de propietarios, sin embargo, cierra estructuralmente ese espacio. Al redistribuir el capital y garantizar la independencia material

universal, transforma la naturaleza misma del poder de negociación. En este contexto, como argumenta Thomas (2017), los incentivos que surgen tienden a ser compensatorios por esfuerzos o desutilidades reales, y no extorsivos. La institución misma genera un ethos de reciprocidad, haciendo innecesario apelar a motivos moralmente exigentes más allá del interés racional dentro de un marco justo. Así, la crítica de Cohen, dirigida a un marco rawlsiano interpretado en el contexto de un capitalismo de bienestar, pierde su objetivo cuando el constructivismo se lleva a su consecuencia institucional plena en la democracia de propietarios. La objeción de Cohen presupone una distinción rígida entre reglas de regulación (supuestamente *fact-dependientes*) y principios fundamentales de justicia (*fact-independent*), que el propio método rawlsiano del equilibrio reflexivo desdibuja al mostrar cómo los principios para la estructura básica necesariamente incorporan, en su justificación, consideraciones sobre su viabilidad y aceptabilidad pública en condiciones reales (Williams, 2024, sec. 7.1). La democracia de propietarios es, en este sentido, la encarnación de un principio de justicia que ya ha internalizado esas consideraciones a través del proceso constructivo.

## *El principio de reciprocidad como mandato constitutivo de la estructura antropológica*

Para Thomas, el principio de reciprocidad es el corazón de la justicia como equidad y la clave para entender la exigencia de una democracia de propietarios, porque es la expresión social de la relación razonable entre personas morales. Rawls lo formula así: “Las instituciones sociales no han de aprovecharse de las contingencias de la dotación natural, ni de la posición social inicial, ni de la buena o mala suerte en el transcurso de la vida, excepto en formas que benefician a todos, incluidos los menos favorecidos” (citado en Thomas, 2017, p. 68). Thomas interpreta esto como un mandato constitutivo para el diseño de la estructura básica. La democracia de propietarios institucionaliza la reciprocidad mediante la difusión universal del capital. Mecanismos como fondos de capital social y herencias universales buscan que “todos los ciudadanos tengan acceso al capital, tanto humano como físico” (Thomas, 2017, p. 20). Esto coloca a los mercados “en un contexto que modela sus resultados como justos de la única manera factible en que esta meta puede ser asegurada” (Thomas, 2017, p. 20). El efecto es crear lo que Rawls llama “estrecha vinculación” (close-knittedness) y “conexión en cadena” (chain-connectedness) entre las expectativas de los diferentes grupos sociales (Thomas, 2017, pp. 66-67). Cuando cada ciudadano es dueño de una porción del capital social, el crecimiento económico beneficia a todos de manera automática y proporcional, realizando el principio

de reciprocidad en la estructura económica misma y generando las condiciones sociales del autorrespeto, un bien primario esencial.

*Respuesta a la crítica de G. A. Cohen: las instituciones como realización del ethos derivado de la antropología*

La crítica de Cohen es central porque pone de relieve la importancia del contexto institucional y la conexión entre estructura básica y acción individual. Cohen acusaba a Rawls de confundir justicia con eficiencia al permitir “incentivos especiales” para los talentosos dentro del principio de diferencia, lo que requeriría un ethos igualitario suplementario en la esfera personal.

La réplica de Thomas, fundamentada en la antropología rawlsiana, es una defensa robusta del institucionalismo constructivista. Sostiene que Cohen “confunde una parte de la visión de Rawls con el todo” (Thomas, 2017, p. 19). En una democracia de propietarios, el principio de diferencia opera en un marco radicalmente distinto donde: (1) todos son propietarios de capital, por lo que el “talento” no puede amenazar con retener un aporte esencial que otros no poseen; (2) el poder político no está capturado por intereses oligárquicos; y (3) la autoestima está protegida por la independencia material generalizada. “Especifica el contexto correcto y la crítica de Cohen desaparece” (Thomas, 2017, p. 19). El diseño institucional de



la democracia de propietarios es lo que asegura que los incentivos funcionen dentro de un “rango permisible” que no dañe la solidaridad. La justicia se realiza a través de la ingeniería institucional, que moldea las opciones e incentivos de todos, haciendo innecesario apelar a un ethos igualitario exigente en la esfera personal. La institución es la materialización del ethos, y ese diseño institucional se deriva de la necesidad de crear condiciones estables para el ejercicio de los poderes morales. El equilibrio reflexivo, al incorporar teorías sobre la motivación y la estabilidad, conduce a preferir diseños institucionales que generen buenos resultados sin depender de virtudes heroicas.

La síntesis liberal-republicana propuesta por Thomas (2017) añade una dimensión crucial al permitir la constitucionalización de las garantías propias de una democracia de propietarios. Esta constitucionalización protege el valor equitativo de las libertades políticas de manera más robusta que la mera “estrategia de aislamiento” y previene eficazmente la deriva oligárquica que tanto preocupaba a Rawls (Thomas, 2017, p. 20). Thomas argumenta que la democracia de propietarios debe entenderse como un “constitucionalismo económico”, con principios estructurales “blindados” en la constitución para asegurar el marco de fondo (*background justice*) de manera perdurable (Thomas, 2017, p. 255). Esto responde al interés de orden superior en la estabilidad y la justicia.

El principio de reciprocidad, tal como es realizado en la democracia de propietarios, es la expresión social de la autonomía. En un marco de cooperación social donde todos poseen una base de independencia material, las interacciones económicas dejan de ser relaciones de dependencia y dominación para convertirse en intercambios entre iguales. La reciprocidad ya no es una mera caridad vertical, sino un reconocimiento horizontal de la dignidad y la contribución de cada participante, un reconocimiento que es constitutivo de la autoestima social.

Frente a la crítica de Malleson (2014) sobre la insuficiente democratización de la inversión y del lugar de trabajo en la democracia de propietarios, una interpretación robusta de los bienes primarios rawlsianos —especialmente “los poderes y prerrogativas de los cargos” y las “bases sociales del autorrespeto”— apunta hacia la necesidad de complementar la democracia de propietarios con instituciones de democracia industrial y control social de la inversión. Esta complementariedad acerca significativamente la democracia de propietarios al socialismo democrático, sin por ello sacrificar su carácter distintivo basado en la propiedad individual dispersa, y puede verse como una especificación adicional que surge del equilibrio reflexivo al considerar el bien primario de las “bases sociales del autorrespeto” en el contexto de la vida laboral.

## **Ideas finales**

A lo largo de este análisis se ha demostrado que la crítica aparentemente fragmentaria de Rawls al capitalismo se unifica bajo un fundamento antropológico coherente, articulado a través del constructivismo kantiano y justificado metodológicamente por el equilibrio reflexivo. Este marco no solo permite derivar la democracia de propietarios como la realización institucional privilegiada de la justicia como equidad, sino que también proporciona recursos para enfrentar críticas canónicas a su proyecto.

Frente al intuicionismo racional o el realismo moral, el constructivismo rawlsiano redefine la objetividad de la justicia como una construcción desde un punto de vista social compartido. Frente a las objeciones de G.A. Cohen, la fundamentación antropológica y el método del equilibrio reflexivo muestran que la incorporación de consideraciones de estabilidad, publicidad y factibilidad no es una concesión espuria, sino parte integral de la justificación de principios destinados a una sociedad pluralista y estable.

Así, la democracia de propietarios emerge no como una mera regla de regulación, sino como la especificación de un principio de justicia que satisface simultáneamente las exigencias normativas de la autonomía y las condiciones prácticas para su realización histórica. Al situar a Rawls en diálogo con estas posiciones críticas, se fortalece la tesis

central: la democracia de propietarios es la consecuencia lógica y más coherente del liberalismo político rawlsiano, ofreciendo un fundamento antropológico y metodológico sólido para una sociedad de ciudadanos verdaderamente libres e iguales.

La democracia de propietarios representa una síntesis virtuosa: es liberal en su protección prioritaria de las libertades básicas y su respeto al pluralismo; es igualitaria en su distribución predistributiva de la riqueza productiva; y es republicana en su garantía de la no-dominación a través de la independencia material. Al hacerlo, institucionaliza las condiciones materiales necesarias para el ejercicio efectivo de la autonomía moral y política, cerrando la brecha entre la teoría ideal y la realidad no ideal, tal como lo exige un equilibrio reflexivo amplio.

En definitiva, la obra de Rawls, leída a través de la lente de su constructivismo kantiano y su método del equilibrio reflexivo, ofrece no solo una crítica poderosa al capitalismo contemporáneo, sino también un horizonte normativo robusto para imaginar una sociedad de ciudadanos verdaderamente libres, iguales y autónomos, donde la propiedad no sea un privilegio sino la condición material universal de la libertad.

## Referencias

- Arnold, N. S. (1994). *The philosophy and economics of market socialism: A critical study*. Oxford University Press.
- Cohen, G. A. (2008). *Rescuing justice and equality*. Harvard University Press.
- Courtland, S. D., Gaus, G., & Schmidtz, D. (2022). Liberalism. En E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Edición de Primavera 2022). Centro de Estudios de Lenguaje e Información, Universidad de Stanford. <https://plato.stanford.edu/archives/spr2022/entries/liberalism/>
- Fuentes, J., & Castillo, J. (2024). Autonomy as political ideal of liberalism. *Telos*, 2024(209), 65–79.
- Guyer, P. (2024). Kantian constructivism I: Rawls. En *Kant's impact on moral philosophy* (pp. 539–577). Oxford University Press.
- Kaufman, A. (2012). Rawls and Kantian constructivism. *Kantian Review*, 17(2), 227–256. <https://doi.org/10.1017/S1369415412000040>
- Knight, C. J. S. (2025). Reflective equilibrium. En E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Edición de Primavera 2025). Centro de Estudios de Lenguaje e Información, Universidad de Stanford.

<https://plato.stanford.edu/archives/spr2025/entries/reflective-equilibrium/>

- Korsgaard, C. M. (2008). Realism and constructivism in twentieth-century moral philosophy. En *The constitution of agency: Essays on practical reason and moral psychology* (pp. 302–326). Oxford University Press.
- Kymlicka, W. (2002). *Contemporary political philosophy: An introduction* (2.<sup>a</sup> ed.). Oxford University Press.
- Malleson, T. (2014). Rawls, property-owning democracy, and democratic socialism. *Journal of Social Philosophy*, 45(2), 228–251. <https://doi.org/10.1111/josp.12052>
- Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.
- O'Neill, M. (2012). Free (and fair) markets without capitalism: Political values, principles of justice, and property-owning democracy. En M. O'Neill & T. Williamson (Eds.), *Property-owning democracy: Rawls and beyond* (pp. 75–100). Wiley-Blackwell. <https://doi.org/10.1002/9781444355192.ch4>
- O'Neill, O. (2015). Constructivism in Rawls and Kant. En *Kant on the construction of reason* (pp. 69–85). Cambridge University Press.
- Rawls, J. (1971). *A theory of justice*. Belknap Press of Harvard University Press.
- Rawls, J. (1980). Kantian constructivism in moral theory. *The Journal of Philosophy*, 77(9), 515–572. <https://doi.org/10.2307/2025790>

- Rawls, J. (1993). *Political liberalism*. Columbia University Press.
- Rawls, J. (2001). *Justice as fairness: A restatement* (E. Kelly, Ed.). Harvard University Press.
- Rawls, J. (2005). *Political liberalism* (Edición expandida). Columbia University Press.
- Raz, J. (1986). *The morality of freedom*. Clarendon Press.
- Ron, A. (2008). Visions of democracy in 'property-owning democracy': Skelton to Rawls and beyond. *History of Political Thought*, 29(1), 168–187.
- Sandel, M. J. (1982). *Liberalism and the limits of justice*. Cambridge University Press.
- Scanlon, T. M. (2003). Rawls on justification. En S. Freeman (Ed.), *The Cambridge companion to Rawls* (pp. 139–167). Cambridge University Press.
- Thomas, A. (2012). Property-owning democracy, liberal republicanism, and the idea of an egalitarian ethos. En M. O'Neill & T. Williamson (Eds.), *Property-owning democracy: Rawls and beyond* (pp. 101–128). Wiley-Blackwell.
- Thomas, A. (2017). *Republic of equals: Predistribution and property-owning democracy*. Oxford University Press.
- Wieland, W. (1992). *Die aristotelische Physik*. Vandenhoeck & Ruprecht.
- Williams, A. (2024). Constructivism in political philosophy. En E. N. Zalta & U. Nodelman (Eds.), *The Stanford*

*Encyclopedia of Philosophy* (Edición de Primavera 2024). Centro de Estudios de Lenguaje e Información, Universidad de Stanford.  
<https://plato.stanford.edu/archives/spr2024/entries/constructivism-political/>



# Gobernanza ética y psicopolítica digital: razón algorítmica, poder y circulación social del conocimiento

*Ethical Governance and Digital Psychopolitics: Algorithmic Reason, Power, and the Social Circulation of Knowledge*

**Paula Andrea Mora Pedreros<sup>77</sup>, Manuel José Santiago Pajajoy<sup>78</sup>**

## Resumen

Este capítulo propone una distinción analítica entre gobernanza ética y psicopolítica digital para comprender los

---

<sup>77</sup> Fundación Universitaria de Popayán. <https://orcid.org/0000-0001-8070-9885>

<sup>78</sup> Fundación Universitaria de Popayán. <https://orcid.org/0000-0003-1730-9218>

dilemas contemporáneos de la circulación social del conocimiento en entornos mediados por algoritmos. La tesis sostiene que la psicopolítica digital describe una lógica de subjetivación y poder que opera mediante personalización, optimización y economía de la atención; en contraste, la gobernanza ética nombra el marco normativo e institucional capaz de orientar, limitar y auditar esa racionalidad algorítmica en nombre de la dignidad, la responsabilidad y la deliberación democrática. Mediante un análisis filosófico-crítico e intertextual, se articulan aportes de Byung-Chul Han (diagnóstico del poder), Adela Cortina (criterio normativo) y Daniel Innerarity (criterio político-institucional), para formular una matriz textual de riesgos y medidas de gobernanza orientadas a mitigar desinformación, injusticias epistémicas y erosión de confianza pública.

**Palabras clave:** Gobernanza ética; psicopolítica digital; razón algorítmica; epistemología social; desinformación.

## **Abstract**

This chapter develops an analytical distinction between ethical governance and digital psychopolitics to address the contemporary challenges of the social circulation of knowledge in algorithmically mediated environments. The central claim is that digital psychopolitics names a logic of subjectivation and power operating through personalization, optimization, and the attention economy; whereas ethical

governance designates the normative and institutional framework able to steer, constrain, and audit algorithmic reason in the name of dignity, responsibility, and democratic deliberation. Through a critical, intertextual philosophical analysis, the chapter brings together Byung-Chul Han's diagnosis of power, Adela Cortina's normative criterion, and Daniel Innerarity's political-institutional perspective, and culminates in a textual matrix linking risks, algorithmic mechanisms, political-epistemic effects, and actionable governance measures aimed at mitigating misinformation, epistemic injustice, and the erosion of public trust.

**Keywords:** Ethical governance; digital psychopolitics; algorithmic reason; social epistemology; misinformation.

## **Introducción**

La inteligencia artificial (IA) se ha consolidado como un componente integral y omnipresente de la sociedad moderna, redefiniendo profundamente la experiencia humana en sus múltiples facetas, desde la comunicación y la educación hasta la salud y el trabajo (Villas Olmeda & Camacho Ibáñez, 2021). Esta transformación tecnológica, sin embargo, trasciende el mero cambio instrumental; representa una reconfiguración ontológica de la realidad social. Como argumenta Daniel Innerarity (2025), la actual datificación no es solo un incremento cuantitativo en el volumen de datos, sino un "cambio cualitativo del proceso de constitución de la

realidad social" (p. 21). Los algoritmos que median nuestras interacciones, decisiones y percepciones no son simples intermediarios pasivos; son arquitectos activos de un nuevo entorno existencial (Sadin, 2018).

Lejos de ser una ruptura radical con el pasado, la emergencia de la IA puede entenderse como la culminación de una larga trayectoria histórica de racionalización y control. Innerarity (2025) sostiene que "la actual digitalización podría entenderse como una continuación intensiva de prácticas de burocratización y racionalización de siglos anteriores" (p. 19). Esta genealogía del poder algorítmico se remonta a la aspiración política fundamental de la modernidad: la búsqueda de la automaticidad para gestionar la complejidad social. Thomas Hobbes, al concebir el Estado como un "automaton", un "hombre artificial", sentó las bases filosóficas para una política entendida como computación, un sistema de cálculo diseñado para imponer orden sobre el caos (Innerarity, 2025, p. 18). Desde la "avalancha de números impresos" del siglo XIX, que permitió a los estados modernos clasificar y administrar a sus poblaciones, hasta la cibernética del siglo XX, la ambición de crear sistemas de gobierno automatizados ha sido una constante (Innerarity, 2025, p. 19). En este sentido, la IA no inventa la voluntad de control, sino que la dota de una eficacia y una escala sin precedentes (Foucault, 1979).

En términos clásicos es repensar lo expuesto por Rawls (1971) en su "Teoría de la justicia" se emplean para contextualizar el debate sobre la equidad y la justicia algorítmica. Su velo de ignorancia es un poderoso dispositivo mental para preguntarnos si los sistemas de IA que diseñamos serían considerados justos por todos los miembros de la sociedad, especialmente los más desfavorecidos. Observaremos más adelante lo promulgado por Kaplan sobre la "automatización de la desigualdad", como un punto en común, al igual que la relación con el concepto de "dignidad humana" expuesto por Kant (1785), al establecer en su ética deontológica como servirá de contrapunto a la lógica utilitarista y consecuencialista que a menudo domina el diseño de algoritmos; esto permitirá argumentar, junto a Cortina (2024), qué es lo que nos define como humanos frente a la máquina.

A partir de lo anterior, este capítulo se organiza en torno a la siguiente pregunta: ¿cómo distinguir la psicopolítica digital —como lógica algorítmica de subjetivación y poder— de una gobernanza ética de lo digital, y qué criterios permiten transformar la razón algorítmica en un objeto de control democrático para proteger la circulación social del conocimiento? La tesis que se defiende sostiene que la psicopolítica digital nombra el mecanismo de control que opera desde la optimización y la transparencia, mientras que la gobernanza ética designa el ensamblaje normativo e institucional orientado a auditar, limitar y democratizar dicha

racionalidad, con base en dignidad, responsabilidad y deliberación pública. En consecuencia, el capítulo aporta:

- Una definición operativa de razón algorítmica como categoría puente entre poder, normatividad y diseño institucional;
- Una articulación crítica Han–Cortina–Innerarity para explicar, evaluar y gobernar los efectos político-epistémicos de la mediación algorítmica;
- Una matriz textual (dimensión; riesgo; mecanismo algorítmico; efecto político-epistémico; principio de gobernanza; medida concreta) que traduce el diagnóstico filosófico en propuestas verificables de intervención.

## **Marco teórico**

### *Definiciones operativas de los conceptos fundamentales*

En este capítulo, la razón algorítmica designa el conjunto de operaciones socio-técnicas —clasificación, predicción, recomendación y modulación de visibilidad— mediante las cuales los sistemas computacionales traducen prácticas sociales a datos, jerarquizan visibilidades y orientan conductas en función de métricas de optimización. Su alcance no se limita a la eficiencia técnica: la razón algorítmica reorganiza las condiciones de credibilidad, autoridad y confianza que sostienen la vida pública del conocimiento,

constituyéndose así en un operador epistemológico-político de primera magnitud (Innerarity, 2025, pp. 16-21). Su límite conceptual reside en que no cubre la totalidad de las prácticas computacionales, sino específicamente aquellas que inciden en la circulación social del conocimiento y en la formación de preferencias, creencias y juicios públicos.

Según Byung-Chul Han (2017), la psicopolítica digital nos aproxima al entendimiento de la lógica de subjetivación y poder propia del capitalismo de plataformas, que opera no mediante la represión externa sino a través de la seducción, la autoexposición y la optimización voluntaria. A diferencia del biopoder foucaultiano, que actúa sobre el cuerpo y la vida biológica, la psicopolítica incide en la psique: coloniza deseos, modula emociones y captura la atención mediante la personalización algorítmica (Han, 2017, p. 24). Su alcance en este capítulo se circunscribe a los mecanismos digitales de gestión subjetiva con incidencia en la esfera pública del conocimiento; quedan fuera de su ámbito los usos terapéuticos o clínicos de la tecnología y las dimensiones de salud mental que, aunque derivadas de la misma lógica, constituyen subdimensiones colaterales y no el núcleo analítico del presente análisis.

Definir la gobernanza ética de lo digital, es una exigencia conceptual ineludible en este contexto; se entiende por tanto, como el ensamblaje de principios normativos, arreglos institucionales y mecanismos de rendición de

cuentas orientados a auditar, limitar y democratizar la razón algorítmica en nombre de la dignidad, la responsabilidad y la deliberación democrática. No se reduce a la ética empresarial voluntaria ni a la mera proclamación de valores; implica obligaciones exigibles, procedimientos de supervisión ciudadana y criterios públicos para evaluar impactos. Su fundamento normativo pivota sobre la ética cordial de Adela Cortina (2024) —que exige condiciones de diálogo veraz y reconocimiento del interlocutor como digno— y sobre el imperativo de autogobierno democrático de Innerarity (2025). El límite de este concepto es que no opera como teoría completa de la regulación tecnológica, sino como criterio normativo-institucional para identificar las condiciones mínimas bajo las cuales la mediación algorítmica resulta compatible con la protección de la circulación social del conocimiento.

El concepto de injusticia epistémica, acuñado por Miranda Fricker (2007), designa el daño infligido a un agente específicamente en su condición de sujeto cognoscente “a wrong done to someone specifically in their capacity as a knower” (p. 1). En su formulación más precisa, adopta dos modalidades principales: la injusticia testimonial, que ocurre cuando un agente recibe un déficit injustificado de credibilidad por causa de prejuicios identitarios; y la injusticia hermenéutica, que emerge cuando una comunidad carece de recursos interpretativos colectivos para comprender y comunicar sus propias experiencias. En el contexto de este



capítulo, la categoría se restringe a sus manifestaciones algorítmicamente mediadas: la asignación sesgada de credibilidad a través de sistemas automatizados de moderación, ranking o visibilidad; y la producción de gaps hermenéuticos cuando ciertos grupos carecen de representación suficiente en los corpus y datasets que entrenan modelos de IA. Su límite operativo excluye las injusticias epistémicas producidas por mecanismos puramente interpersonales o no digitales, cuyo análisis requiere metodologías diferenciadas.

### *Una crítica filosófica a la digitalización*

Para desentrañar las complejidades de esta nueva era, este artículo se apoya en un marco teórico multidisciplinario, con la filosofía de Byung-Chul Han como eje central. El pensamiento de Han ofrece un diagnóstico incisivo de las patologías de la sociedad digital, articulado a través de conceptos clave que se han vuelto indispensables para su crítica (Han, 2010, 2013, 2017). La "sociedad de la transparencia" describe un panóptico digital donde la coacción a la visibilidad total elimina la privacidad y la negatividad, sometiendo al individuo a un control exhaustivo (Han, 2013). La "hiperconexión" disuelve las estructuras comunitarias en un "enjambre" de individuos aislados, bombardeados por un exceso de información que atrofia la capacidad de reflexión (Han, 2017). Finalmente, la "sociedad del cansancio" o de la positividad revela un régimen de

rendimiento neoliberal donde el sujeto, convertido en empresario de sí mismo, se autoexplota hasta el agotamiento (Han, 2010). Estos conceptos no solo describen fenómenos aislados, sino que juntos componen un sistema de dominación que Han (2017) denomina "psicopolítica", un poder inteligente que no reprime, sino que seduce y optimiza desde dentro.

Este marco se enriquece con la perspectiva de Adela Cortina (2024), quien advierte sobre un "eclipse de la razón comunicativa" (p. 145). En una sociedad tecnologizada, la proliferación de canales de conexión no garantiza una comunicación más profunda; por el contrario, puede degradarla a un mero intercambio estratégico de información, vaciándola de veracidad y de su potencial para construir un entendimiento mutuo. Cortina (2024) señala que la hiperconectividad no produce, por sí misma, comunicación auténtica ni deliberación (p. 26) y que este eclipse amenaza el fundamento mismo de la deliberación democrática, ya que "es justamente la razón comunicativa la que está quedando eclipsada, cuando es indispensable para construir una vida humana justa y feliz, y, por supuesto, sociedades democráticas" (Cortina, 2024, p. 27).

Por todo lo anterior, se integra la "crítica de la razón algorítmica" de Daniel Innerarity (2025), quien propone ir más allá de la simple exhortación ética para examinar las condiciones estructurales y las relaciones de poder

incrustadas en la propia arquitectura tecnológica. Su enfoque nos obliga a cuestionar la supuesta neutralidad de los algoritmos y a politizar el debate sobre la digitalización, reconociendo que cada artefacto tecnológico es una forma de "filosofía encubierta" que moldea nuestras prácticas sociales y políticas (Agre, 1997, p. 240, citado en Innerarity, 2025, p. 17).

Es aquí donde emerge la categoría de la razón algorítmica al convertirse en puente que articula tres niveles: (i) el nivel psicopolítico, donde Han describe la lógica contemporánea de subjetivación/poder vinculada a transparencia, autoexposición y control; (ii) el nivel normativo, donde Cortina aporta el criterio ético de dignidad, responsabilidad y ética pública para distinguir conexión de comunicación y exigir condiciones de diálogo; y (iii) el nivel político-institucional, donde Innerarity (2025) aporta el criterio de gobernanza democrática de lo digital, orientado a la supervisión ciudadana, la rendición de cuentas y el diseño de controles institucionales. El aporte del ensamblaje es epistemológico-social: permite responder al problema contemporáneo de circulación social del conocimiento, en tanto traduce mecanismos algorítmicos en efectos político-epistémicos (confianza, credibilidad, pluralismo informativo, injusticias epistémicas) y, por tanto, en principios y medidas concretas de gobernanza.

## *Epistemología social de la razón algorítmica*

La relevancia del concepto de razón algorítmica avizorada por (Innerarity, 2025) no se reduce a la eficiencia técnica, sino a la reorganización de las condiciones de credibilidad, autoridad y confianza que sostienen la vida pública del conocimiento. Por ello, el problema no es únicamente “qué hace” un sistema, sino qué tipo de ecología epistémica produce: qué voces eleva; cuáles margina; qué afirmaciones vuelven verosímiles; y qué desacuerdos vuelve improbables.

Este desplazamiento exige pasar de una epistemología centrada en el sujeto aislado a una epistemología social atenta a instituciones, mediaciones y prácticas colectivas. Goldman advierte que “Epistemology has had a strongly individualist orientation, at least since Descartes” (Goldman, 2009, p. 1); sin embargo, en entornos digitales la formación de creencias, la evaluación de fuentes y la estabilización de consensos dependen de infraestructuras socio-técnicas que distribuyen, amplifican o silencian información. En consecuencia, la pregunta epistémica se vuelve institucional: quién define los criterios de relevancia; cómo se asigna credibilidad; bajo qué reglas se corrigen errores; y qué incentivos moldean el ecosistema informativo.

La categoría de injusticia epistémica esbozada con anterioridad aporta un lente decisivo para leer los efectos de

la razón algorítmica sobre la circulación social del conocimiento. Fricker (2007) acerta en el señalamiento sobre este fenómeno, porque muestra que los daños epistémicos no son abstractos: se materializan cuando ciertos grupos reciben déficits sistemáticos de credibilidad; cuando las plataformas convierten rasgos identitarios en señales de sospecha; o cuando una comunidad carece de recursos interpretativos para nombrar, comprender y denunciar aquello que experimenta. La automatización de decisiones de visibilidad (qué se recomienda; qué se modera; qué se “rankea”) puede, así, reproducir o intensificar asimetrías de voz bajo la apariencia de neutralidad computacional.

La psicopolítica digital, en clave haniana, permite comprender cómo esa infraestructura epistémica se acopla a dinámicas de poder que operan desde el interior de la subjetividad. Cuando la visibilidad se convierte en mandato, la exposición deja de ser un acto comunicativo y se transforma en dispositivo de captura de datos y de gobierno de la conducta. Han lo formula con contundencia: “Esta vigilancia total degrada la ‘sociedad transparente’ hasta convertirla en una inhumana sociedad del control” (Han, 2017, p. 90). Epistemológicamente, ello supone un desplazamiento de la confianza hacia la trazabilidad: se privilegia lo medible; lo registrable; lo cuantificable; y se debilitan las mediaciones deliberativas mediante las cuales una comunidad negocia sentidos, valida testimonios y corrige creencias.

En este punto se vuelve indispensable la distinción entre psicopolítica digital y gobernanza ética. La primera nombra el mecanismo de subjetivación/poder; la segunda introduce el criterio normativo e institucional para poner límites, generar obligaciones y diseñar garantías. Cortina aporta una advertencia que funciona como umbral crítico: “conectarse no es comunicarse, no es dialogar” (Cortina, 2024, p. 27). La razón algorítmica produce conexión, pero no asegura diálogo; no garantiza reciprocidad; ni responsabilidad por la veracidad; ni reconocimiento del otro como interlocutor digno. En términos de epistemología social, el déficit no es solo informativo, sino cívico: sin condiciones para el diálogo, la esfera pública se vuelve vulnerable a la desinformación, a la manipulación afectiva y a la erosión del criterio compartido.

Innerarity aporta el criterio político-institucional al situar la IA en el horizonte del autogobierno democrático. Su tesis es tajante: “si la ciudadanía no puede supervisar las decisiones públicas, entonces no podemos llamar a eso autogobierno del pueblo” (Innerarity, 2024, párr. 3). Desde la epistemología social, esta afirmación implica que la gobernanza de lo digital no es un añadido externo a la técnica; es condición de posibilidad de una esfera pública epistémicamente saludable. La supervisión democrática requiere traducir la razón algorítmica en objeto de control: auditorías; reglas de rendición de cuentas; vías de reclamación; y mecanismos de corrección cuando el sistema

amplifica falsedades, asigna credibilidad de modo sesgado o segmenta de forma opaca la conversación pública.

De ahí que una gobernanza ética de la razón algorítmica deba expresarse en medidas concretas orientadas a la circulación social del conocimiento: transparencia significativa (explicabilidad útil para usuarios y control público, no solo para cumplimiento formal); responsabilidad por impactos epistémicos (cómo un sistema afecta confianza, reputación y visibilidad de grupos); garantías de pluralismo informativo (criterios de diversidad en recomendación y mitigación de burbujas); y alfabetización crítica en IA, entendida como competencia cívica para evaluar fuentes, rastrear incentivos y disputar narrativas. En conjunto, la epistemología social permite formular el núcleo del argumento: la promesa tecnológica solo es defendible si se acompaña de principios y arreglos institucionales que protejan el testimonio, la deliberación y la justicia epistémica en la era de la razón algorítmica.

**Tabla 1**

*Matriz textual de análisis para la razón algorítmica: dimensiones, riesgos, efectos político-epistémicos y medidas de gobernanza ética*

| Dimensión                       | Riesgo                                                         | Mecanismo algorítmico                                                                            | Efecto político-epistémico                                                                                                 | Principio de gobernanza                             | Medida concreta                                                                                                                                          |
|---------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Visibilidad / transparencia     | Vigilancia y autoexposición; ampliación de asimetrías de poder | Trazabilidad; perfilamiento; ranking de presencia; inferencia de atributos                       | Autocensura; disciplinamiento; inhibición del testimonio; erosión de confianza                                             | Privacidad; proporcionalidad; minimización de datos | Límites estrictos a recolección y retención; Evaluación de Impacto (DPIA/AIA) ; auditorías de vigilancia; rutas de reclamación                           |
| Personalización / recomendación | Burbujas informativas; polarización; captura de agenda         | Sistemas de recomendación; filtrado colaborativo; microsegmentación; optimización por engagement | Fragmentación epistémica; cierre deliberativo; debilitamiento del desacuerdo productivo; pérdida de pluralismo informativo | Pluralismo informativo; transparencia significativa | Métricas de diversidad y exposición cruzada; auditorías de amplificación; controles sobre microtargeting; reportes públicos de criterios de priorización |



| <b>Dimensión</b>                      | <b>Riesgo</b>                                              | <b>Mecanismo algorítmico</b>                                                                | <b>Efecto político-epistémico</b>                                                                     | <b>Principio de Medida</b>                             | <b>Medida concreta</b>                                                                                                                   |
|---------------------------------------|------------------------------------------------------------|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|--------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Optimización / eficiencia             | Sustitución del juicio por métricas; desresponsabilización | Scoring; KPIs; automatización de flujos; predicción de desempeño                            | Reducción del criterio; dependencia de indicadores; desplazamiento de la argumentación por “puntajes” | Rendición de cuentas; revisión humana significativa    | Trazabilidad decisional; obligación de motivación; revisión humana en decisiones de alto impacto; mecanismos de apelación y corrección   |
| Moderación / credibilidad             | Injusticia epistémica; censura errónea; sobre-etiquetado   | Modelos de detección; umbrales de moderación; etiquetas; shadow banning                     | Déficits/excesos de credibilidad por grupo; silenciamiento; degradación de confianza en reglas        | Justicia epistémica; no discriminación; debido proceso | Pruebas de sesgo; transparencia de normas y criterios; auditorías externas; notificación y explicación; canales de reclamación efectivos |
| Generación / síntesis (IA generativa) | Autoridad epistémica indebida; contaminación informacional | Generación probabilística; síntesis sin trazabilidad; alucinación; reescritura automatizada | Degradación de veracidad; confusión fuente-autoridad; erosión de confianza pública                    | Integridad informativa; trazabilidad; verificación     | Etiquetado y divulgación de uso; trazabilidad de fuentes; verificación y citación obligatoria en contextos                               |

| <b>Dimensión</b>                   | <b>Riesgo</b>                                              | <b>Mecanismo algorítmico</b>                                                   | <b>Efecto político-epistémico</b>                                                        | <b>Principio de gobernanza</b>                   | <b>Medida concreta</b>                                                                                                                   |
|------------------------------------|------------------------------------------------------------|--------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|--------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
|                                    |                                                            |                                                                                |                                                                                          |                                                  | académicos; políticas de uso responsable                                                                                                 |
| Datos / gobernanza de dataset      | Reproducción de sesgos; extracción no consentida; opacidad | Recolección masiva; scraping; datasets no documentados; correlaciones espurias | Sesgos sistemáticos; pérdida de legitimidad; desconfianza en conocimiento “automatizado” | Legalidad; calidad; accountability de datos      | Documentación (datasheets); gobernanza de consentimiento; control de procedencia; auditoría de calidad y sesgo; límites de reutilización |
| Seguridad / integridad del sistema | Manipulación; ataques; adulteración de outputs             | Data poisoning; prompt injection; model inversion; adversarial examples        | Distorsión de evidencia; vulneración de confianza; amplificación de falsedad             | Seguridad por diseño; resiliencia                | Red teaming; pruebas adversariales; monitoreo de deriva; controles de acceso; protocolos de respuesta a incidentes                       |
| Deliberación / esfera pública      | Captura afectiva; manipulación política;                   | Nudges algorítmicos; diseño persuasivo; microtargeting                         | Manipulación de preferencias; debilitamiento del debate                                  | Deliberación democrática; transparencia política | Restricciones a microtargeting político; archivos                                                                                        |

| Dimensión | Riesgo                | Mecanismo algorítmico   | Efecto político-epistémico                   | Principio de Medida gobernanza concreta                                                                      |
|-----------|-----------------------|-------------------------|----------------------------------------------|--------------------------------------------------------------------------------------------------------------|
|           | deterioro democrático | político; dark patterns | público; desdemocratización del conocimiento | públicos de anuncios; auditorías de influencia; estándares de explicabilidad para sistemas de impacto cívico |

*Nota.* Elaboración propia.

### *Planteamiento del problema y pregunta de investigación*

La convergencia de estas tecnologías y dinámicas sociales plantea una serie de desafíos urgentes que requieren un análisis filosófico profundo (Sadin, 2018). La usabilidad de la IA, entendida no solo como su eficiencia técnica sino como su capacidad para integrarse y amplificar las lógicas de la transparencia, la hiperconexión y la positividad, genera un nuevo ecosistema humano-tecnológico con profundas implicaciones éticas. La investigación se articula en torno a la siguiente pregunta central, refinada a partir del texto base: *¿Cómo opera la razón algorítmica como mecanismo de psicopolítica digital que reconfigura la credibilidad y la confianza pública, y qué criterios de gobernanza ética —*

*normativos e institucionales— permiten democratizar su incidencia en la circulación social del conocimiento?*

La relevancia de esta cuestión es crucial. Abordarla implica la necesidad imperativa de desarrollar marcos teóricos y regulatorios que puedan guiar el desarrollo y la implementación de la IA de una manera que sea verdaderamente humana (Villas Olmeda & Camacho Ibáñez, 2021). Se trata de asegurar que el progreso tecnológico no se produzca a costa de valores fundamentales como la autonomía, la privacidad, la dignidad y el bienestar psicológico. La filosofía crítica, por tanto, no es un mero ejercicio académico, sino una herramienta indispensable para imaginar y construir un futuro en el que la tecnología sirva al florecimiento humano en lugar de a su empobrecimiento (Pariser, 2011).

Desde esta articulación, el problema deja de ser exclusivamente tecnológico o moral; se vuelve epistemológico-social: la razón algorítmica distribuye visibilidad y credibilidad, reordena la confianza y condiciona la producción pública de verdad. Por ello, la distinción gobernanza ética/psicopolítica digital solo es consistente si se traduce en criterios para gobernar los efectos político-epistémicos de la mediación algorítmica (pluralismo informativo, justicia epistémica, posibilidad de deliberación y corrección pública del error).

## Metodología

Para abordar la compleja pregunta de investigación, se ha empleado un enfoque metodológico cualitativo, centrado en una revisión documental exhaustiva y un análisis hermenéutico (Denzin & Lincoln, 2011). Este método permite una comprensión profunda y holística de las implicaciones éticas y sociales de la IA, yendo más allá de la mera descripción de fenómenos para interpretar su significado en el contexto de las transformaciones socioculturales contemporáneas. Es pertinente precisar el registro argumentativo del presente trabajo: su pretensión no es explicativa-causal ni empírica, sino conceptual-normativa. Las inferencias que se formulan a lo largo del texto son inferencias teóricas derivadas del análisis filosófico-hermenéutico de las fuentes primarias; no constituyen constataciones empíricas ni generalizaciones estadísticas. Cuando en los resultados se afirma que la razón algorítmica “produce” determinados efectos o que ciertos mecanismos “generan” consecuencias político-epistémicas, debe entenderse que tales afirmaciones expresan relaciones conceptuales y tendencias identificadas por el marco teórico adoptado, susceptibles de verificación empírica en investigaciones posteriores. La revisión de la literatura, como proceso sistemático, consiste en “detectar, consultar y obtener la bibliografía y otros materiales útiles para los propósitos del estudio” (Gómez-Luna et al., 2014, p. 430), con el fin de establecer un marco teórico sólido y comprender el

estado actual del conocimiento en el campo (Hernández-Sampieri et al., 2014).

El componente hermenéutico de esta metodología es fundamental. No se trata de una simple recopilación de datos, sino de un ejercicio interpretativo que busca desentrañar las complejidades conceptuales y las tensiones dialécticas presentes en la intersección de la IA, la filosofía y la crítica social (Bryman & Bell, 2015). Se busca comprender cómo los conceptos desarrollados por los pensadores seleccionados dialogan entre sí, se complementan o se contradicen, para construir una síntesis analítica coherente y matizada. Este enfoque permite explorar las capas de significado subyacentes en los discursos sobre la tecnología, revelando los supuestos, valores e ideologías que a menudo permanecen implícitos.

El corpus de esta investigación ha sido cuidadosamente seleccionado para proporcionar una perspectiva multifacética y rigurosa sobre el tema. Se estructura en torno a un núcleo de fuentes primarias, complementado por un conjunto de textos interdisciplinarios que amplían y contextualizan el análisis.

- Fuentes Primarias: El eje del análisis lo constituyen las obras fundamentales de Byung-Chul Han, específicamente *La sociedad de la transparencia* (2013), *El enjambre* (2017), *La sociedad del cansancio* (2010) y *Psicopolítica* (2017). Estos textos proporcionan las

categorías analíticas centrales que estructuran toda la investigación.

- Fuentes Complementarias: Para enriquecer y dar profundidad al marco de Han, se ha integrado un conjunto de obras de referencia que abordan el fenómeno de la IA desde diferentes ángulos:
  - Contexto Histórico-Político: *Una teoría crítica de la inteligencia artificial* de Daniel Innerarity (2025) es crucial para entender la razón algorítmica como una evolución de las prácticas de burocratización y gobernanza.
  - Dimensión Humanista y Comunicativa: *¿Ética o ideología de la inteligencia artificial?* de Adela Cortina (2024) aporta una perspectiva indispensable sobre la definición de lo humano frente a la máquina y la crisis de la democracia deliberativa.
  - Implicaciones Socioeconómicas: *Abstenerse humanos* de Jerry Kaplan (2016) ofrece un análisis detallado y concreto de las consecuencias de la automatización en el trabajo, la riqueza y la desigualdad.
  - Perspectivas Críticas y Psicológicas Adicionales: Se incorporan las reflexiones de Éric Sadin (2018) sobre la "administración digital del mundo", de Stanislas Dehaene (2014) sobre la conciencia y el cerebro, y de

Marian Rojas Estapé (2024) sobre el impacto de la tecnología en la salud mental.

- Fuentes para la Discusión: En la sección final del artículo, se recurrirá a dos textos clave para proponer marcos de respuesta a los desafíos identificados: el *Decálogo por la transparencia y ética del uso de la inteligencia artificial en la producción académica* de Diego F. Craig (2024) y el estudio sobre la figura del “Eticista de la IA” de Cocchiaro et al. (2024). Estos documentos, de carácter propositivo-práctico y utilizados aquí como referencias contextuales —no como fundamento conceptual del marco teórico—, permiten pasar del diagnóstico filosófico a la propuesta normativa y profesional. Conviene señalar que Cocchiaro et al. (2024) corresponde a un manuscrito no publicado y, en consecuencia, sus propuestas se incorporan con la cautela metodológica propia del material gris, sin atribuirles el peso epistemológico de la literatura académica revisada por pares.

El proceso de análisis se ha desarrollado en tres fases interconectadas, diseñadas para construir una argumentación sólida y progresiva (Creswell & Plano Clark, 2011):

1. Extracción y Definición Conceptual: En la primera fase, se han extraído y definido rigurosamente los conceptos centrales del pensamiento de Byung-Chul



Han: la sociedad de la transparencia, la hiperconexión y el enjambre, y la sociedad de la positividad y el cansancio (Han, 2010, 2013, 2017).

2. **Análisis Categórico y Síntesis Intertextual:** En la segunda fase, estos conceptos han sido utilizados como categorías analíticas para organizar y examinar los argumentos, evidencias y ejemplos presentados por los demás autores del corpus. Por ejemplo, el análisis de Kaplan (2016) sobre la vigilancia del consumidor en Amazon se ha interpretado a través de la lente de la "sociedad de la transparencia" de Han (2013). De manera similar, la crisis de la democracia deliberativa descrita por Cortina (2024) se ha conectado con el concepto de "enjambre" (Han, 2017).
3. **Construcción de una Crítica Multidimensional:** La tercera y última fase ha consistido en sintetizar los hallazgos de este análisis intertextual para construir una crítica multidimensional y coherente sobre el impacto de la IA. Este procedimiento permite demostrar que los desafíos de la IA no son meramente técnicos, sino que están profundamente arraigados en las patologías sociales y políticas de nuestra época, tal como las diagnostica la filosofía crítica contemporánea (Innerarity, 2025).

## **Resultados hermenéuticos**

### *El diagnóstico de la sociedad digital a través de la IA*

El análisis de la literatura revela que la inteligencia artificial no opera en un vacío, sino que se inserta y amplifica las dinámicas estructurales de la sociedad digital contemporánea. Utilizando el marco conceptual de Byung-Chul Han, es posible diagnosticar cómo la IA actúa como un catalizador de tres patologías centrales: la tiranía de la visibilidad a través de la transparencia algorítmica, la erosión del ser causada por la hiperconexión y el eclipse de la razón comunicativa, y la dictadura de la eficiencia que culmina en la automatización de la desigualdad.

Leído desde epistemología social, este régimen de visibilidad no solo produce control; también altera criterios de credibilidad: lo visible se confunde con lo verdadero y lo cuantificable desplaza la confianza. Mecanismo algorítmico: ranking de presencia, perfilamiento y trazabilidad; efecto político-epistémico: autocensura y debilitamiento del testimonio; principio de gobernanza: privacidad y proporcionalidad; medida concreta: límites de recolección, auditorías de vigilancia y rutas de reclamación ante decisiones automatizadas de exposición.

## *La tiranía de la visibilidad: transparencia algorítmica y control psicopolítico*

Byung-Chul Han (2013) describe la "sociedad de la transparencia" no como un ideal de apertura democrática, sino como una "coacción sistémica" que convierte cada aspecto de la vida en información visible, cuantificable y, por tanto, controlable. Es un régimen donde la intimidad, el secreto y la negatividad —aquello que se resiste a la comunicación inmediata— son eliminados en favor de una exposición total. Han (2013) afirma que “la transparencia se ha convertido en una coacción sistémica. Todo se convierte en transparente y visible para poder ser controlado y monitoreado” (p. 12). La inteligencia artificial emerge como el motor tecnológico por excelencia de este panóptico digital, dotándolo de una capacidad de vigilancia y análisis sin precedentes (Foucault, 1979).

Esta dinámica se manifiesta de forma palpable en lo que Jesús M. Flores-Vivar y Francisco J. García-Peñalvo (2023) denominan "vigilancia algorítmica". Los sistemas de IA, a través de la recopilación masiva y el procesamiento de datos, pueden monitorear el comportamiento humano en tiempo real, generando perfiles detallados que no solo registran nuestras acciones pasadas, sino que predicen e influyen en nuestras decisiones futuras. Este nivel de control plantea cuestiones éticas fundamentales sobre la autonomía y la privacidad. Como advierten los autores, “la vigilancia

algorítmica debe ser regulada para asegurar que los derechos de privacidad y autonomía sean respetados. La transparencia debe ser balanceada con la protección de los datos personales” (Flores-Vivar & García-Peñalvo, 2023, p. 52).

El análisis de Jerry Kaplan (2016) en *Abstenerse humanos* ofrece una ilustración económica de este mecanismo de control. Los "intelectos sintéticos" que operan en el comercio electrónico, como los de Amazon, o en las finanzas, a través de la negociación de alta frecuencia (HFT), se fundamentan en la extracción y explotación masiva de datos. Estos sistemas no solo venden productos o gestionan activos; su función principal es construir perfiles exhaustivos de los individuos para predecir y manipular su comportamiento. Cada clic, cada búsqueda, cada compra se convierte en un dato que alimenta un modelo diseñado para optimizar la persuasión, transformando a los ciudadanos en consumidores predecibles y gestionables (Kaplan, 2016).

Lo que emerge de la convergencia de estas perspectivas es una fusión inquietante entre las lógicas históricas de control estatal y las nuevas formas de control corporativo. Daniel Innerarity (2025) traza una genealogía que conecta la aspiración del Estado moderno a la calculabilidad y la automatización para gestionar la complejidad social —la "avalancha de números impresos" del siglo XIX— con la razón algorítmica actual. Por su parte, Kaplan (2016) muestra cómo las corporaciones utilizan la IA para alcanzar un nivel de

control sin precedentes sobre los mercados y los consumidores. El concepto de "psicopolítica" de Han (2017) unifica estos dos fenómenos. El poder en la era digital ya no opera principalmente a través de la represión externa (disciplina), sino a través de la seducción, la optimización y la auto-vigilancia (control). Es un poder que nos incita a exponernos voluntariamente, a cuantificarnos y a optimizarnos, sirviendo así tanto a los intereses del neoliberalismo de mercado como a los del control estatal. La IA, por tanto, no crea un tipo de poder fundamentalmente nuevo, sino que perfecciona y fusiona el poder soberano del Estado y el poder de mercado de las corporaciones en un único y sofisticado sistema de gobernanza psicopolítica, donde la transparencia total es la herramienta universal de control.

*La erosión del ser: hiperconexión, fatiga informativa y el eclipse de la razón comunicativa*

El segundo pilar de la crítica de Han (2017) es el concepto de "enjambre", que describe la estructura social de la era digital. A diferencia de la "masa" de la era industrial, que poseía una identidad y un alma colectiva capaz de acción política conjunta, el enjambre es una "agregación de individuos aislados". Estos individuos están hiperconectados, pero carecen de un "nosotros" cohesivo. La comunicación digital, según Han (2017), "genera una especie de enjambre que se mueve en una dirección caótica, sin una narrativa

coherente, afectando nuestra capacidad de concentración y reflexión” (p. 45). Esta sobrecarga informativa y la exigencia de disponibilidad constante conducen a una profunda fatiga psicológica.

Las consecuencias de la hiperconexión en la salud mental y la atención cognitiva constituyen una subdimensión colateral pertinente, aunque no el núcleo central del presente análisis. Desde la perspectiva adoptada en este capítulo, lo relevante epistemológicamente no es el agotamiento individual en sí mismo, sino su impacto en la capacidad deliberativa colectiva: un sujeto fragmentado y sobreinformado pierde progresivamente la aptitud para razonar, evaluar fuentes y participar en la conversación pública con criterio. En ese sentido, los análisis de Rojas Estapé (2024) sobre la ansiedad asociada a la hiperconexión y de Dehaene (2014) sobre el deterioro atencional funcionan, en el marco de este trabajo, como evidencias contextuales que ilustran un efecto derivado de la razón algorítmica sobre la ecología cognitiva del agente epistémico, no como fundamentos teóricos del argumento principal. La consecuencia directamente relevante para la epistemología social es la siguiente: el enjambre digital no sólo fragmenta la comunidad; erosiona las condiciones psíquicas mínimas para la escucha, la argumentación reflexiva y el disenso productivo que toda deliberación democrática requiere.

En este contexto, el concepto de Adela Cortina (2024) del "eclipse de la razón comunicativa" adquiere una relevancia crucial. La hiperconexión del enjambre es, precisamente, el mecanismo que produce este eclipse. La comunicación, en su sentido más profundo, es un acto dialógico orientado al entendimiento mutuo, basado en la veracidad y el respeto por el interlocutor (Cortina, 2024). Sin embargo, en el enjambre digital, se degrada a un mero intercambio de información, a menudo estratégico y performativo. Como señala Cortina (2024), la tragedia de nuestra era es que, a pesar de tener más canales de conexión que nunca, hemos olvidado cómo comunicarnos genuinamente. La velocidad y la superficialidad de la interacción digital socavan la paciencia, la escucha y la argumentación reflexiva que son indispensables para la construcción de un consenso democrático.

Se revela aquí una paradoja fundamental de la sociedad digital: existe una relación causal inversa entre el nivel de *conectividad* tecnológica y la fortaleza de la *comunidad* social y política. La inteligencia artificial, al optimizar la conectividad individualizada a través de algoritmos de personalización y filtros burbuja, contribuye activamente al desmantelamiento del espacio público compartido (Pariser, 2011). La arquitectura de la IA actual, centrada en la recomendación y la optimización de la experiencia del usuario individual, no es políticamente neutral. Su diseño intrínseco fomenta el aislamiento

narcisista del enjambre, donde cada individuo habita una realidad algorítmicamente curada para él. Esto se conecta con la "despolitización algorítmica" de la que habla Innerarity (2025), donde las decisiones complejas y controvertidas se presentan como soluciones técnicas objetivas, eludiendo el debate y el pluralismo que son la esencia de la política democrática. Por lo tanto, la amenaza que la IA representa para la democracia deliberativa no es accidental, sino estructural. Al promover la conexión sin comunidad, la IA provoca el eclipse de la razón comunicativa, erosionando las bases mismas del autogobierno democrático.

En clave de circulación social del conocimiento, la sustitución del juicio por optimización afecta la posibilidad misma de deliberar: la conversación pública se reorganiza por incentivos de atención, no por fuerza de razones. Mecanismo algorítmico: recomendación y microsegmentación; efecto político-epistémico: fragmentación epistémica, pérdida de pluralismo y cierre del desacuerdo productivo; principio de gobernanza: pluralismo informativo y explicabilidad significativa; medida concreta: métricas de diversidad en recomendación, auditorías de ampliación y obligaciones de transparencia sobre criterios de priorización.



## *La dictadura de la eficiencia: positividad, autoexplotación y la automatización de la desigualdad*

La tercera patología diagnosticada por Han (2010) es la "sociedad del cansancio", un régimen dominado por el imperativo de la positividad. En esta sociedad, el lema no es la obediencia, sino la iniciativa; el verbo modal no es "deber", sino "poder". El sujeto de rendimiento, impulsado por la creencia de que "nada es imposible", se convierte en "un empresario de sí mismo, autoexplotándose hasta el punto de la extenuación" (Han, 2010, p. 22). La violencia ya no proviene de un agente externo que reprime, sino del propio individuo que se exige un rendimiento ilimitado, llevando a enfermedades neuronales como la depresión, el trastorno por déficit de atención con hiperactividad (TDAH) y el síndrome de desgaste profesional (burnout).

La inteligencia artificial actúa como catalizador de esta cultura de la hiperproductividad, aunque es preciso enmarcar este fenómeno correctamente en el eje analítico central del presente capítulo: el vínculo entre poder, algoritmos y circulación social del conocimiento. La aceleración del rendimiento y sus efectos sobre el bienestar individual son, desde este marco, subdimensiones colaterales de un problema estructural más profundo: la eficiencia algorítmica desplaza el juicio razonado por la métrica optimizada, erosionando así la responsabilidad epistémica en la esfera pública. Como señala Sadin (2018), la "administración digital

del mundo” reduce a los individuos a componentes de un sistema productivo, pero su consecuencia más relevante para la epistemología social no es el agotamiento psíquico en sí, sino la normalización de la obediencia a indicadores en detrimento de la deliberación argumentada. Esta sustitución del criterio por la métrica es lo que debilita la posibilidad misma de producción colectiva y crítica de conocimiento.

El análisis socioeconómico de Jerry Kaplan (2016) en *Abstenerse humanos* proporciona la manifestación material de esta crítica filosófica. Kaplan argumenta que la automatización impulsada por la IA está reconfigurando la lucha de clases. La clásica confrontación marxista entre capital y trabajo se transforma en una lucha de "activos contra personas" (Kaplan, 2016, p. 17). En una economía optimizada por la IA, los "intelectos sintéticos" y los "falsos trabajadores" son inherentemente más eficientes que los humanos. La búsqueda incesante de la máxima productividad lleva a la conclusión lógica de que los humanos, como el eslabón más débil y menos eficiente de la cadena productiva, deben abstenerse ("Humans need not apply"). Esto no solo provoca un desplazamiento masivo de empleos, sino que, como afirma Kaplan (2016), "el verdadero problema es que los ricos necesitarán pocas personas, o ninguna, que trabajen para ellos" (p. 17).

Este proceso puede entenderse como la culminación de la lógica neoliberal. La IA no solo transforma el mercado

laboral; completa un ciclo ideológico que Han (2010) había diagnosticado. La "sociedad del cansancio" representa la fase de entrenamiento en la que los seres humanos internalizan y normalizan la lógica de la eficiencia y la auto-optimización. La automatización, tal como la describe Kaplan (2016), es la fase final de este proceso, en la que esa misma lógica se externaliza y se materializa en máquinas que nos superan en rendimiento. El resultado es la "automatización de la desigualdad" (Brynjolfsson & McAfee, 2014), un sistema en el que los propietarios del capital (los activos) ya no necesitan del trabajo (las personas), llevando la lógica de la explotación a su conclusión definitiva: la irrelevancia económica y social del trabajador. La promesa de liberación del trabajo a través de la tecnología se invierte así en una amenaza de exclusión masiva del sistema productivo, un futuro en el que la eficiencia algorítmica triunfa sobre la necesidad humana.

La eficiencia algorítmica, sin gobernanza, no solo precariza el trabajo; también precariza la responsabilidad epistémica: se normaliza obedecer métricas sin justificar razones. Mecanismo algorítmico: scoring, automatización de desempeño y evaluación continua; efecto político-epistémico: desresponsabilización y reducción del criterio; principio de gobernanza: rendición de cuentas y revisión humana significativa; medida concreta: trazabilidad decisonal, obligación de motivación, y mecanismos de apelación ante decisiones automatizadas con impacto en derechos.

## Discusión e implicaciones

Los resultados del análisis convergen en una conclusión central: la inteligencia artificial, en su forma actual de desarrollo e implementación, no es una fuerza exógena que crea nuevas patologías sociales desde cero. Más bien, actúa como un poderoso catalizador y acelerador de las crisis ya latentes en la sociedad digital contemporánea, diagnosticadas por pensadores como Byung-Chul Han (2010, 2013, 2017). La IA intensifica la vigilancia de la sociedad de la transparencia, fragmenta la comunidad en un enjambre hiperconectado pero aislado, y lleva la lógica autoexplotadora de la sociedad del cansancio a su culminación a través de la automatización. El resultado es la formación de un "enjambre algorítmico": una colectividad de individuos gestionados por sistemas opacos, atomizados por la personalización y agotados por un imperativo de rendimiento que, en última instancia, amenaza con volverlos obsoletos (Kaplan, 2016; Pariser, 2011). Estos desafíos no son meros efectos secundarios o fallos técnicos que puedan corregirse con ajustes algorítmicos; son problemas estructurales que exigen una respuesta ética, política y social a múltiples niveles.

*Hacia una gobernanza ética y una resistencia humana: un marco de respuesta multinivel*

Frente a un diagnóstico tan complejo, las soluciones simplistas como una moratoria tecnológica generalizada o la

mera apelación a códigos éticos voluntarios resultan insuficientes (Innerarity, 2025). Se requiere un marco de respuesta robusto y multinivel que aborde el problema desde la práctica individual hasta la estructura política global. Este marco puede articularse en tres niveles de intervención: micro (individual y profesional), meso (institucional y organizacional) y macro (político y social).

En el nivel más fundamental (micro), la resistencia al determinismo algorítmico comienza con la práctica individual y profesional. El *Decálogo por la transparencia y ética* de Diego F. Craig (2024) ofrece un marco de referencia práctica —con carácter contextual y orientado a la aplicación, no a la fundamentación conceptual del capítulo— para guiar la acción de investigadores, educadores, estudiantes y desarrolladores. Principios como el “Uso Crítico”, que “implica una evaluación meticulosa y reflexiva de cómo estas tecnologías se aplican en la investigación y la educación” (Craig, 2024, p. 21); la “Conciencia de Sesgos”, que obliga a desvelar y corregir las predisposiciones inherentes a los modelos; y la “Claridad en la Autoría”, que demanda una transparencia total sobre el grado de intervención de la IA, constituyen formas operativas de resistencia epistémica y ética. Al adoptar estas prácticas, los individuos pueden contrarrestar activamente la opacidad, el automatismo y la falta de responsabilidad que caracterizan a muchos sistemas de IA, reafirmando la primacía del juicio humano crítico (Craig, 2024).

Desde luego, la responsabilidad individual, aunque necesaria, no es suficiente. Es crucial integrar la reflexión ética dentro de las estructuras institucionales que desarrollan y despliegan la IA. En este nivel, la propuesta de Cocchiaro et al. (2024) de formalizar la figura del “Eticista de la IA” ilustra, con carácter orientativo y a título de referencia contextual — dado que se trata de un manuscrito no publicado—, cómo podría institucionalizarse la reflexión ética en las organizaciones. Este nuevo rol profesional no debe ser un mero ejercicio de “ethics washing”, sino una función empoderada con la misión de ser “facilitador, investigador y educador” dentro de las organizaciones (Cocchiaro et al., 2024, p. 14). El Eticista de la IA estaría encargado de realizar evaluaciones éticas, gestionar los conflictos de valores, analizar políticas, mediar en disputas y, fundamentalmente, capacitar al personal para fomentar una cultura de responsabilidad. La institucionalización de esta figura, cuya pertinencia normativa y conceptual se fundamenta principalmente en la gobernanza democrática propuesta por Innerarity (2025) y en los principios de responsabilidad de Cortina (2024), apunta a convertir la reflexión crítica en parte integral del ciclo de vida de la tecnología.

### *Nivel macro: la acción política y social.*

Finalmente, los desafíos más profundos de la IA solo pueden abordarse en la arena política y social. Siguiendo a Daniel Innerarity (2025), es imperativo llevar a cabo una

"politización" de la tecnología, rechazando el tecnosolucionismo que presenta las decisiones algorítmicas como neutrales y objetivas. El desarrollo y la gobernanza de la IA deben ser objeto de un debate democrático amplio y pluralista, donde se cuestionen sus supuestos y se delibere sobre sus fines. Este llamado a la politización se conecta directamente con la defensa de Adela Cortina (2024) de la "razón comunicativa". Revitalizar el espacio público para un diálogo genuino, veraz y orientado al entendimiento es la forma más potente de resistencia democrática frente a la fragmentación del enjambre y el eclipse de la deliberación. Se trata de que la sociedad, a través de sus instituciones democráticas, recupere el control sobre su futuro tecnológico, en lugar de aceptarlo pasivamente como un destino inevitable.

La siguiente tabla sintetiza este marco de respuesta multinivel, mostrando cómo diferentes propuestas teóricas y prácticas pueden integrarse en una estrategia holística para la gobernanza ética de la IA.

**Tabla 2**

*Un Marco Multinivel para la Gobernanza Ética de la IA*

| Nivel de Intervención   | Objetivo Principal                                                       | Actores Clave                                             | Propuesta Concreta                                                                               | Fuente Principal                  |
|-------------------------|--------------------------------------------------------------------------|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------------------------------|
| Micro (Individual)      | Fomentar la práctica ética y el pensamiento crítico.                     | Investigadores, educadores, estudiantes, desarrolladores. | Adopción de un decálogo de buenas prácticas (transparencia, conciencia de sesgos, uso crítico).  | Craig (2024)                      |
| Meso (Institucional)    | Integrar la reflexión ética en los procesos de desarrollo y despliegue.  | Empresas, universidades, organizaciones.                  | Creación y empoderamiento del rol profesional del "Ético de la IA".                              | Cocchiaro et al. (2024)           |
| Macro (Político-Social) | Someter la tecnología al control democrático y a los valores humanistas. | Ciudadanía, parlamentos, sociedad civil.                  | "Politización" de la IA y revitalización de la "razón comunicativa" a través del debate público. | Innerarity (2025), Cortina (2024) |



*Nota.* Elaboración propia.

### *Limitaciones y futuras líneas de investigación*

En síntesis, la psicopolítica digital permite describir el modo en que la razón algorítmica gobierna subjetividades y reconfigura la credibilidad pública; la gobernanza ética, por su parte, nombra la condición normativa e institucional para someter ese poder a control democrático. La distinción no es meramente terminológica: organiza el capítulo como una respuesta a la circulación social del conocimiento, al vincular mecanismos algorítmicos con efectos político-epistémicos y, finalmente, con medidas de gobernanza. Con ello, el texto pasa del diagnóstico a la propuesta: delimita riesgos; fija principios; y ofrece un repertorio de intervenciones auditables orientadas a proteger deliberación, testimonio y confianza pública.

Es importante reconocer que este estudio, por su naturaleza, es primordialmente teórico y filosófico. Su principal contribución radica en la construcción de un marco conceptual robusto para interpretar los desafíos de la IA, pero no ofrece una validación empírica de sus tesis. Esta limitación, sin embargo, abre la puerta a futuras líneas de investigación que puedan poner a prueba, contextualizar y

refinar los hallazgos aquí presentados, tal como se sugería en la metodología original del texto base.

Las futuras investigaciones deberían orientarse a cerrar la brecha entre la crítica filosófica y la realidad empírica. Se proponen las siguientes direcciones:

- Estudios de caso sobre el impacto de la IA en el bienestar laboral: Investigaciones empíricas en sectores específicos (como la logística, el servicio al cliente o el desarrollo de software) podrían medir cuantitativa y cualitativamente el impacto de la automatización y la gestión algorítmica en la salud mental de los trabajadores, validando la conexión entre los análisis de Han (2010) y Kaplan (2016).
- Análisis etnográficos de la comunicación en plataformas digitales: Estudios cualitativos profundos sobre cómo las plataformas de redes sociales, con sus algoritmos de recomendación, reconfiguran el debate público, la formación de identidades colectivas y la posibilidad de una comunidad política. Esto permitiría observar en la práctica el "eclipse de la razón comunicativa" de Cortina (2024) y la formación del "enjambre" de Han (2017).

- Investigaciones comparativas de marcos regulatorios: Un análisis sistemático de las diferentes aproximaciones regulatorias a la IA (como la Ley de IA de la Unión Europea, el enfoque de mercado de EE. UU. o el modelo de control estatal de China) para evaluar su eficacia real a la hora de abordar los problemas de control, transparencia, sesgo y responsabilidad identificados por Innerarity (2025) y Flores-Vivar y García-Peñalvo (2023).

Al integrar estas perspectivas críticas y avanzar hacia una investigación empíricamente fundamentada, es posible avanzar hacia un futuro donde la tecnología y la ética coexistan armoniosamente, garantizando que los beneficios de la IA sean accesibles para todos, sin comprometer los valores humanos fundamentales y, sobre todo, potenciando las habilidades que nos definen como humanos (Cortina, 2024).

## **Conclusiones**

Este recorrido exploratorio, anclado en la crítica filosófica, permite articular una serie de conclusiones fundamentales sobre los desafíos de la inteligencia artificial en la sociedad contemporánea. Lejos de ser una mera herramienta tecnológica, la IA se revela como una fuerza que

moldea profundamente la experiencia humana, intensificando dinámicas sociales y de poder preexistentes.

En primer lugar, desde el marco hermenéutico adoptado, cabe inferir que la transparencia algorítmica, tal como opera en la arquitectura actual de los sistemas de IA, no realiza el ideal deliberativo de apertura pública, sino que puede configurarse —según el diagnóstico de Han (2013) leído a través de la epistemología social— como un mecanismo de control que debilita la privacidad, inhibe el testimonio y erosiona la confianza. Esta inferencia, de naturaleza conceptual, sugiere que la regulación debe distinguir entre transparencia operativa —la que sirve al control ciudadano de decisiones públicas— y transparencia extractiva —la que refuerza asimetrías de poder entre plataformas y usuarios—. El criterio normativo, derivado de Cortina (2024), es que solo la primera es compatible con las condiciones del diálogo democrático.

En segundo lugar, la hiperconexión gestionada algorítmicamente produce, desde el marco conceptual de Han (2017) y la epistemología social de Goldman (2009), una fragmentación de las condiciones cognitivas e institucionales para la deliberación pública. Esta inferencia teórica no equivale a constatar empíricamente que todos los usuarios de plataformas digitales sufren deterioro deliberativo; significa que la arquitectura de recomendación y microsegmentación, orientada por incentivos de engagement, produce presiones

estructurales contrarias a la escucha, la argumentación reflexiva y el disenso productivo. Los impactos sobre el bienestar psicológico individual —ansiedad, fatiga atencional, dispersión— son efectos colaterales de esta misma arquitectura, pero su relevancia en este capítulo es instrumental: evidencian la degradación del agente epistémico que participa en la esfera pública. El criterio normativo central sigue siendo el “eclipse de la razón comunicativa” (Cortina, 2024, p. 27) como amenaza a la democracia deliberativa, no el bienestar individual en sí mismo.

En tercer lugar, la sustitución del juicio por la métrica —lógica central de la “sociedad del cansancio” (Han, 2010) amplificada por la optimización algorítmica— produce, desde una perspectiva normativa, una desresponsabilización epistémica: los actores institucionales y profesionales delegan en indicadores automatizados decisiones que exigirían deliberación, justificación pública y rendición de cuentas. Esta inferencia teórica es relevante para la gobernanza porque identifica el riesgo normativo de aceptar como neutras decisiones que tienen efectos distribucionales sobre credibilidad, visibilidad y acceso al conocimiento. La respuesta institucional que se deriva no es solo la protección del bienestar laboral —aunque ésta sea valiosa en sí misma— sino la exigencia de trazabilidad decisional, obligación de motivación y mecanismos de apelación ante decisiones automatizadas con impacto en derechos.

Del análisis precedente se derivan cuatro prioridades políticas explícitas para la gobernanza ética de la razón algorítmica, junto a los criterios que permiten evaluar su impacto efectivo sobre la circulación social del conocimiento. Estas prioridades no son meras recomendaciones de buenas prácticas: operan como condiciones normativas mínimas para que la mediación algorítmica resulte compatible con la dignidad, la responsabilidad y la deliberación democrática (Cortina, 2024; Innerarity, 2025). La primera prioridad es la implantación de auditorías algorítmicas obligatorias y accesibles públicamente para sistemas de alto impacto en la esfera pública del conocimiento; el criterio de evaluación correspondiente es la reducción medible de sesgos sistemáticos y de asimetrías de credibilidad por razón de grupo, en términos de Fricker (2007), la reducción de injusticia testimonial mediada algorítmicamente. La segunda prioridad es la transparencia significativa: no la mera divulgación formal de criterios técnicos, sino la explicabilidad útil para usuarios, periodistas, reguladores y organizaciones de la sociedad civil; el criterio de evaluación es el incremento del control ciudadano efectivo sobre decisiones automatizadas con impacto en visibilidad, credibilidad y acceso a información pública, medible a través de indicadores de pluralismo informativo (diversidad de fuentes amplificadas, exposición cruzada a perspectivas, reducción de burbujas de filtración). La tercera prioridad es la introducción de límites exigibles al microtargeting político y a los dark

patterns de persuasión en contextos electorales y de deliberación pública; el criterio de evaluación es el mantenimiento —o restauración— de las condiciones mínimas para la formación autónoma de preferencias por parte de la ciudadanía, medible a través de indicadores de confianza pública en instituciones democráticas e integridad electoral. La cuarta prioridad es la implantación de debido proceso en moderación y curaduría automatizada de contenidos, que incluya notificación, explicación y vías efectivas de reclamación; el criterio de evaluación es la disminución de errores sistemáticos de silenciamiento que constituyen injusticia epistémica, evaluable mediante auditorías externas de equidad y análisis diferenciados por grupo. En conjunto, estas cuatro prioridades traducen la distinción gobernanza ética/psicopolítica digital en un programa de intervención auditable, cuyo éxito debe medirse en términos de tres dimensiones complementarias: incremento del pluralismo informativo efectivo, sostenimiento o recuperación de la confianza pública en instituciones y fuentes de conocimiento, y reducción cuantificable de la injusticia epistémica mediada algorítmicamente.

Finalmente, la integración de estas perspectivas críticas nos lleva a una conclusión propositiva. El futuro de la IA no es un destino tecnológico predeterminado, sino un campo de decisión ética y política (Innerarity, 2025). Para navegarlo con rigor, es necesario abandonar la ingenuidad

tecnológica y adoptar un enfoque crítico, normativo y proactivo, que vincule el diagnóstico filosófico con prioridades verificables de intervención institucional, tal como se ha planteado en el presente capítulo. La inteligencia artificial, en última instancia, es una herramienta de multiplicación: su valor depende enteramente de lo que decidamos multiplicar. Si partimos de la indiferencia, la explotación y el control, el resultado será una distopía eficiente. Si, por el contrario, partimos de la dignidad, la deliberación y el autogobierno democrático, la IA puede convertirse en un instrumento para potenciar lo mejor de nuestras capacidades epistémicas y políticas colectivas (Cortina, 2024; Innerarity, 2025). La pregunta que nos interpela no es qué puede hacer la IA, sino qué clase de comunidad epistémica queremos construir en la era de su existencia.

## Referencias

- Bryman, A., & Bell, E. (2015). *Business research methods* (4.<sup>a</sup> ed.). Oxford University Press.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Cocchiaro, M. Z., Morley, J., Novelli, C., Panai, E., Tartaro, A., & Floridi, L. (2024). *Who is an AI ethicist? An empirical study of expertise, skills, and profiles to build a competency framework* [Manuscrito en



- preparación]. Oxford Internet Institute, University of Oxford.
- Cortina, A. (2024). *¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnologizada*. Paidós.
- Craig, D. F. (2024). *Decálogo por la transparencia y ética del uso de la inteligencia artificial en la producción académica*.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2.<sup>a</sup> ed.). SAGE Publications.
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin UK.
- Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE handbook of qualitative research* (4.<sup>a</sup> ed.). SAGE Publications.
- Etapé, M. R. (2024). *Recupera tu mente, reconquista tu vida*. Grijalbo.
- Flores-Vivar, J. M., & García-Peñalvo, F. J. (2023). Aspectos éticos de la vigilancia y el control algorítmico en la sociedad de la transparencia. *Revista de Ética y Tecnología*, 15(2), 45–62.
- Foucault, M. (1979). *Discipline and punish: The birth of the prison*. Vintage.
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

- Goldman, A. I. (2009). Social epistemology: Theory and applications. *Royal Institute of Philosophy Supplement*, 64. Cambridge University Press.
- Gómez-Luna, E., Jara-Rojas, R., & Reyes-Lagos, J. J. (2014). Revisión de literatura en investigación cualitativa y cuantitativa. *Revista Colombiana de Ciencias Sociales*, 5(2), 429–444.
- Han, B.-C. (2010). *La sociedad del cansancio*. Herder.
- Han, B.-C. (2013). *La sociedad de la transparencia*. Herder.
- Han, B.-C. (2014). *En el enjambre* (R. Gabás, Trad.). Herder. (Obra original publicada en 2013)
- Han, B.-C. (2017). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder*. Herder.
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la investigación* (6.<sup>a</sup> ed.). McGraw-Hill.
- Innerarity, D. (2024). Si la ciudadanía no puede supervisar las decisiones públicas, no podemos llamar a eso autogobierno del pueblo. *El País*. <https://elpais.com/tecnologia/2024/daniel-innerarity-ia-autogobierno>
- Innerarity, D. (2025). *Una teoría crítica de la inteligencia artificial*. Galaxia Gutenberg.
- Kant, I. (1785). *Fundamentación de la metafísica de las costumbres*. Alianza Editorial.

- Kaplan, J. (2016). *Abstenerse humanos: Guía para la riqueza y el trabajo en la era de la inteligencia artificial*. Teell Editorial.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.
- Rawls, J. (1971). *A theory of justice*. Harvard University Press.
- Sadin, E. (2018). *La humanidad aumentada: La administración digital del mundo*. Editorial Anagrama.
- Villas Olmeda, M., & Camacho Ibáñez, J. (2021). *Manual de ética aplicada en inteligencia artificial*. Aranzadi.

# **Trade, Security, and Selective Alignment: Process Tracing the Peru- United States Trade Promotion Agreement and the Reorientation of Peruvian Foreign Policy**

*Comercio, seguridad y alineación selectiva: un  
análisis del proceso del Acuerdo de Promoción  
Comercial entre Perú y Estados Unidos y la  
reorientación de la política exterior peruana*

**Anthony Rolando Medina Rivas Plata<sup>79</sup>, Aaron Ubaldo  
del Carpio Retamozo<sup>80</sup>**

---

<sup>79</sup> Universidad Católica de Santa María. <https://orcid.org/0000-0002-5118-5477>

<sup>80</sup> Universidad Católica de Santa María. <https://orcid.org/0000-0003-1927-416X>

## **Abstract**

This article argues that the Peru-United States Trade Promotion Agreement (Peru-US FTA), in addition to being a trade instrument, also functioned as a strategic alignment mechanism that consolidated Peru's open-market model and reoriented Peruvian foreign policy within an ideologically contested Andean order at the beginning of the 21st century. Using process tracing, the paper reconstructs the causal chain linking the temporary nature of the ATPA/ATPDEA tariff preferences, the US strategy of "competitive liberalization" following the stagnation of the FTAA, and the domestic institutional lock-in generated by a rules-intensive treaty. The empirical basis combines official documents from both countries with secondary literature on trade agreements, Peruvian foreign policy, and Andean regional politics. The findings show that the uncertainty stemming from the temporary preferences strengthened Peruvian trade technocrats and export coalitions that demanded a binding agreement that would generate additional costs for a potential regulatory reversal. At the same time, Washington framed the negotiations in terms that were simultaneously commercial, democratic, and anti-drug, treating Peru as a reliable partner in a context of regional polarization. We conclude by noting that Peru actively sought the benefits of closer ties with the United States while accepting a deeper integration into a regional order favored by Washington.

**Keywords:** Peru; United States; foreign policy; trade agreements; process tracing; international political economy.

## **Resumen**

Este artículo sostiene que el Acuerdo de Promoción Comercial Perú-Estados Unidos (TLC Perú-EE.UU.) además de instrumento comercial funcionó también un dispositivo de alineamiento estratégico que consolidó el modelo aperturista peruano y reorientó la política exterior peruana dentro de un orden andino disputado ideológicamente a inicios del siglo XXI. Mediante un mecanismo explicativo de rastreo de procesos (*process tracing*), el trabajo reconstruye la cadena causal que vinculó la temporalidad de las preferencias arancelarias ATPA/ATPDEA, la estrategia estadounidense de 'liberalización competitiva' tras el estancamiento del ALCA y el *lock-in* institucional doméstico generado por un tratado intensivo en reglas. La base empírica combina documentos oficiales de ambos países con literatura secundaria sobre tratados comerciales, política exterior peruana y política regional andina. Los hallazgos muestran que la inseguridad derivada de las preferencias temporales fortaleció a los tecnócratas comerciales peruanos y a las coaliciones exportadoras que demandaban un acuerdo vinculante que genere costos extras a una potencial reversión regulatoria; a la vez que Washington enmarcó la negociación en términos simultáneamente comerciales, democráticos y antidrogas, tratando al Perú como un socio confiable en un contexto de

polarización regional. Concluimos señalando que el Perú buscó activamente los beneficios de un mayor acercamiento a Estados Unidos mientras aceptaba una inserción más densa en un orden regional preferido por Washington.

**Palabras clave:** Perú; Estados Unidos; política exterior; tratados comerciales; process tracing; economía política internacional.

## **Introducción**

Trade agreements are often analyzed in terms of export growth, investment attraction, and tariff reduction. Yet research on North–South preferential trade agreements has shown that their effects extend beyond commerce alone. Such agreements can shape policy autonomy, affect bargaining relationships, and influence the broader diplomatic choices available to participating states (Manger, 2009; Shadlen, 2005, 2008). This broader political significance has been particularly evident in Latin America, where regional and bilateral trade agreements have often been negotiated not only in response to economic opportunities, but also in the context of wider debates over development models, external dependence, and hemispheric order (Phillips, 2005; Quiliconi & Wise, 2009).

The Peru–United States Trade Promotion Agreement provides a useful case for examining how economic and

strategic considerations can converge. Peru negotiated with Washington while rebuilding democratic legitimacy after the collapse of the Fujimori-Montesinos regime, preserving an outward-oriented economic model, and navigating a region increasingly divided between market-friendly and post-neoliberal political projects. At the same time, key export sectors had already expanded under the unilateral preferences granted through ATPA and ATPDEA. Those preferences created important opportunities, but they were temporary, conditional, and subject to shifts in U.S. domestic politics. Even before the treaty was signed, then, Peru's integration into global markets was shaped by a broader political relationship with the United States (Kat, 2023; St John, 2017; Villarreal, 2006).

Existing scholarship has addressed important parts of this story, though often from separate angles. Legal and economic studies have examined the treaty's provisions, trade flows, labor effects, and broader welfare outcomes (Alarco Tosoni, 2017; Arrieta Padilla, 2017; Canessa Montejo, 2020; Cantuarias Salaverry & Raygada, 2008). Work in foreign policy has instead highlighted Peru's diplomatic pragmatism, the consolidation of free trade within the state apparatus, and the later emergence of Pacific-oriented regionalism (McClintock & Vallas, 2010; Ramírez Bullón, 2018; St John, 2017). What remains less explored is how these dynamics connect on how a trade agreement may help shape longer-term foreign-policy orientation. The question, then, is not



only whether exports expanded after 2009, but how a deep agreement may have reinforced a particular external alignment and made departures from that orientation more difficult. The Peruvian case is also conceptually useful because it falls between two familiar, but incomplete, interpretations. One treats Peru as a straightforward success story of trade opening and treaty-based credibility. Another portrays it largely as a dependent actor adapting to agendas set elsewhere. Neither fully captures the case. Peru exercised agency, but it did so within a markedly asymmetric environment. This makes the PTPA a useful case for thinking about how middle and smaller powers use deep economic agreements to pursue international insertion under constraints they do not fully control. What makes the case particularly revealing is the interaction of domestic initiative, external pressures, and regional geopolitical tensions.

This article asks through what mechanisms did the Peru–U.S. agreement shape Peruvian foreign policy? It argues that the PTPA can be understood, at least in part, as a mechanism of selective alignment rather than simply as a commercial accord. The agreement did not create a pro-U.S. orientation from scratch; Peru under Alejandro Toledo had already moved toward pragmatic economic openness and cooperative relations with Washington. Rather, the PTPA reinforced that orientation through three interrelated mechanisms. First, uncertainty surrounding ATPA/ATPDEA strengthened the position of trade technocrats and export-

oriented coalitions seeking a binding framework for market access and credibility. Second, the United States pursued the agreement as part of a broader hemispheric strategy that tied trade to democracy promotion, counternarcotics cooperation, and the search for reliable partners in a polarized Andean region. Third, because the treaty was a deep agreement (with rules on investment, intellectual property, procurement, labor, environment, and dispute settlement) it generated institutional lock-in within Peru, raising both the domestic and diplomatic costs of reversal.

The article contributes to debates on trade and foreign policy in three ways. Empirically, it reconstructs the negotiation and implementation process through process tracing rather than treating the agreement as background context. Theoretically, it brings scholarship on trade agreements into conversation with debates on alignment, suggesting that deep economic institutions can, under some conditions, operate as instruments of selective bandwagoning in asymmetric settings. Substantively, it situates the PTPA within Peru's broader trajectory of globalization. The treaty generated tangible gains in trade and export diversification, but its most enduring significance lies in the way it anchored Peru's preferred model of open regionalism and consolidated Peru's differentiated position vis-à-vis the Bolivarian axis.

The sections that follow develop this argument in stages. The next section reviews the relevant literature and

theoretical expectations. A subsequent section explains the research design, followed by a reconstruction of the historical background from ATPA to the PTPA. The later sections trace the proposed mechanisms and their foreign-policy and economic effects, before a final section considers the broader implications for the study of trade and foreign policy in Latin America.

## **Literature Review and Theoretical Framework**

The broader literature on North-South trade agreements helps explain why such treaties can carry political significance even when negotiated in the language of efficiency and market access. For Bütte and Milner (2008), trade agreements do more than lower tariffs; they also provide credible commitments that reduce policy uncertainty and reassure foreign investors. This is particularly important for developing countries seeking capital, long-term export planning, and external validation of domestic reforms. Manger (2009) likewise shows that North-South preferential agreements are structured by asymmetry. Developed states often seek deeper commitments on services, investment, regulatory governance, and property rights, while developing countries accept those terms in exchange for market access, credibility, and at times geopolitical support. In this sense, the agreement becomes a governance device rather than simply a customs instrument.

A second strand of scholarship emphasizes the developmental and distributive consequences of deep integration. Shadlen (2005) shows that agreements with the United States and other advanced economies often exchange policy autonomy for secure access to markets. For developing countries, this can be a difficult bargain. With greater certainty for exporters and investors, but tighter constraints on industrial policy, intellectual property regulation, and domestic discretion. In a broader analysis of the Americas, Shadlen (2008) argues that the spread of bilateral and regional agreements cannot be understood apart from global competition, hemispheric asymmetry, and the strategic use of bilateralism by the United States. In this view, FTAs are embedded in a broader contest over who shapes the rules of globalization in the hemisphere.

This broader geopolitical dimension became especially salient in the Americas after the stagnation of the Free Trade Area of the Americas. Phillips (2005) argues that U.S. economic governance in the hemisphere has long relied on institutional forms that appear technical but carry deeply political effects. After the FTAA lost momentum, Washington increasingly turned to bilateral and subregional agreements as a more feasible and controllable route to market opening and regulatory convergence. Quiliconi and Wise (2009) describe this as part of a strategy of competitive liberalization that advanced asymmetric FTAs with willing or strategically valuable partners. From this perspective, the Peru–U.S.

agreement was not an isolated bilateral initiative, but part of a broader shift from ambitious hemispheric institutional design toward targeted agreements with states prepared to lock in a liberal economic model.

Domestic politics provides a third analytical layer. Putnam's (1988) two-level games framework reminds us that international bargains succeed only when negotiators can assemble domestic coalitions capable of ratification and implementation. This is particularly relevant for agreements such as the PTPA, whose effects ran through multiple ministries, legal sectors, social groups, and export interests. Von Braun's (2012) work on intellectual-property bargaining in the U.S.–Peru and U.S.–Colombia FTAs shows that trade diplomacy in the Andes was shaped by organized domestic interests, transnational lobbying, and the uneven capacity of local actors to intervene in technical negotiations. Cantuarias Salaverry and Raygada (2008) likewise stress that the agreement's content and implementation depended on complex institutional adaptation within Peru. Any convincing explanation must therefore connect external asymmetry with internal coalition-building rather than privilege one over the other.

The Peru-specific literature supplies important parts of this story, but rarely reconstructs the causal chain in full. McClintock and Vallas (2010) emphasize the close working relationship that developed between Washington and Lima in

the 2000s. St John (2017) argues that Peruvian foreign policy after 2000 exhibited continuity and pragmatism, combining market opening, active diplomacy, and a preference for workable ties with the United States. Ramírez Bullón (2018), meanwhile, shows that the PTPA and the later Pacific Alliance marked critical moments in the legitimization of free trade within Peruvian foreign policy. Kat (2023) comes closest to the argument advanced here by interpreting the U.S.–Peru FTA as a case of competitive liberalization that helped contain post-neoliberal regionalism. What remains underdeveloped, however, is a structured account of how that political effect emerged and why the treaty produced durable diplomatic consequences.

Scholarship on Andean free trade agreements also suggests that much of their political significance lies beneath what often appears to be technical negotiation. Issues such as customs procedures, sanitary regulations, procurement rules, or services commitments may seem far removed from foreign policy, yet they can generate enduring relationships of dependence, compliance, and bureaucratic coordination with geopolitical implications. Seen this way, the Peru–U.S. agreement is not simply a case about trade law or export outcomes. It is also a case about how deep agreements embed states in legal and institutional orders that shape, and sometimes narrow, their diplomatic options. The PTPA is especially revealing in this regard because its dense legal architecture emerged in a highly polarized regional setting.

To account for those foreign-policy effects, this article draws on theories of alignment in realist thought. Walt (1987) showed that states respond not only to concentrations of power but to perceived threats, while Schweller (1994) emphasized that weaker states may align upward when doing so yields material advantage or strategic protection. The Peru–U.S. relationship was not a military alliance, and the PTPA should not be treated as the equivalent of a security pact. Still, the broader logic is relevant. Economic institutions can also serve as vehicles through which weaker states draw closer to dominant powers in pursuit of access, credibility, and political support.

This is where selective bandwagoning becomes useful. Peru did not simply defer to U.S. preferences, nor did it forfeit autonomy in foreign affairs. It expanded ties with Asia, later concluded an FTA with China, and maintained pragmatic relations with neighboring governments despite recurring tensions. Yet amid the regional political divisions of the mid-2000s, Peru did not position itself between competing projects. It moved more decisively toward a Pacific-oriented and rule-based order that overlapped with U.S. preferences. The PTPA helped entrench that orientation by embedding it in law, bureaucratic practice, and domestic coalitions. In that sense, the agreement did not merely reflect an existing alignment; it helped reproduce the conditions that sustained it.

This interpretation also sharpens the distinction between alignment and continuity. An alternative reading might suggest that Peru was already liberal, pragmatic, and broadly aligned with Washington, making the treaty little more than an expression of an existing foreign-policy orientation. That argument captures part of the story. Yet continuity alone cannot account for why the agreement acquired major political significance, generated regional effects, or made later governments bear substantial costs in moving away from the model it consolidated. The claim here is not that the PTPA triggered an ideological transformation, but that it converted a broad orientation into a more deeply institutionalized and binding order.

More precisely, the argument centers on a three-step causal mechanism. First, the temporary and conditional character of ATPA/ATPDEA preferences pushed Peruvian state actors and export coalitions toward seeking a permanent framework with Washington. Second, the United States pursued Peru within a broader regional strategy linking trade, investment, democratic governance, and counternarcotics cooperation in the Andes. Third, because the treaty that emerged was unusually far-reaching, its implementation produced institutional lock-in within Peru, fostering domestic constituencies and policy routines that helped reproduce diplomatic continuity and an outward-looking regional orientation.



This interpretation must also be assessed against three competing explanations. One treats the treaty principally as an economic instrument aimed at export growth and investment attraction. While that view can account for commercial outcomes, it does not explain a broader reorientation of foreign policy. A second stresses continuity, arguing that Peru's pragmatic diplomacy predated the agreement and likely would have endured in its absence. A third privileges ideology, attributing Peru's posture largely to anti-Chavista preferences among political leaders, particularly Alan García, rather than to the treaty itself. Process tracing is particularly useful for weighing these alternatives because it makes it possible to examine the sequencing, content, and evidentiary strength of within-case observations instead of relying on ex post correlations.

## **Research Design**

This study uses explaining-outcome process tracing within a single-case design. Its objective is not to assess the average impact of free trade agreements on foreign policy, but to reconstruct the particular causal pathway through which the Peru–U.S. treaty contributed to a reorientation of Peru's external strategy. Process tracing is especially useful when the outcome under study is historically specific, when competing explanations overlap, and when analysis must move beyond sequence alone toward identifying causal mechanisms (Beach & Pedersen, 2013; Bennett & Checkel, 2015; Collier, 2011;

George & Bennett, 2005). That is the challenge posed by this case. The central question is not simply whether the PTPA preceded changes in Peruvian diplomacy, but how the agreement connected domestic coalitions, U.S. strategic objectives, and regional politics in ways that made alignment more durable over time.

The outcome to be explained is the consolidation of Peru's selective alignment with Washington within a contested Andean regional setting from the decision to negotiate in 2003 through the early implementation years after 2009. The analysis follows this process from the final phase of the ATPA/ATPDEA preference regime through negotiation, ratification, and post-entry-into-force effects, while also considering economic legacies visible by the agreement's fifteenth and sixteenth years. The case is analytically significant for four reasons. Peru was highly vulnerable to the expiration of unilateral preferences; negotiations unfolded during a hemispheric transition following the breakdown of the FTAA; the treaty constituted a deep agreement that demanded extensive legal adaptation at the domestic level; and Peru's subsequent regional behavior allows an assessment of whether those commitments endured.

The process-tracing strategy is structured around a sequence of theoretically grounded observations rather than a straightforward before-and-after comparison. The analysis

begins by examining whether Peru exhibited an endogenous demand for permanence beyond ATPA/ATPDEA. It then considers whether U.S. official discourse framed the treaty as serving multiple strategic purposes. A third step explores whether the agreement's depth generated legal and bureaucratic lock-in within Peru. Finally, the analysis assesses whether regional behavior after ratification aligns more closely with selective alignment than with a merely economic adjustment. This sequencing matters because each stage poses a distinct inferential question and taken together, they allow competing explanations to be evaluated rather than simply narrated.

The evidentiary foundation draws on both primary and secondary sources. Primary materials include the 2003 USTR notification to Congress announcing negotiations with the Andean countries; official USTR and MINCETUR accounts of the agreement; treaty texts and implementation records; Congressional Research Service reports issued during negotiation and ratification; the U.S. International Trade Commission's assessment of the treaty's expected effects; and official trade documentation from both governments. These materials are complemented by scholarly work on Peruvian foreign policy, U.S. hemispheric strategy, Andean regionalism, labor and regulatory effects, and trade outcomes. Together, they enable triangulation, with official documents showing how governments framed their objectives, while

scholarship situates those claims within wider political and institutional dynamics.

Following Bennett and Checkel (2015), the analysis assesses observations according to their inferential weight. Some forms of evidence function as hoop tests, to know if they are necessary for the argument to remain credible, even if they are not independently conclusive. Evidence that Peru actively pursued a permanent treaty because ATPA/ATPDEA preferences were temporary and politically uncertain falls within this category. Without such evidence, the claim that Peru acted as an active demander of the treaty would be considerably weakened. Other observations operate as smoking-gun tests. Most important is the explicit U.S. framing of the Andean FTA as an instrument not only for trade and investment, but also for democracy, stability, and counternarcotics cooperation. Because this appears in official U.S. documents prior to ratification, it carries substantial probative force against a purely economic explanation. Finally, post-ratification developments (continued counternarcotics cooperation, erosion of Andean bloc cohesion, the expansion of Pacific-oriented integration, and the persistence of the treaty across governments) provide cumulative straw-in-the-wind evidence consistent with the hypothesized mechanism.

The method also has important limits. First, the article does not contend that the treaty shaped every aspect of

Peruvian foreign policy. Peru continued to diversify its external relations and remained, in many respects, a pragmatic rather than doctrinaire actor. Second, publicly available sources reveal official positions and institutional effects far more readily than they capture confidential bargaining or private diplomatic exchanges. Interviews with negotiators and access to diplomatic archives would provide additional leverage for strengthening the argument. Third, economic outcomes are incorporated into the explanation, but they are not treated as the sole basis for judging significance. The broader claim is precisely that the foreign-policy significance of the PTPA extends beyond trade creation or GDP growth alone.

**Table 1**

*Mechanism, evidence, and inferential leverage*

| Mechanism step                                           | Observable implications                                                                                 | Principal sources                     | Inferential leverage                        |
|----------------------------------------------------------|---------------------------------------------------------------------------------------------------------|---------------------------------------|---------------------------------------------|
| Preference insecurity and Peruvian demand for permanence | Peruvian officials describe ATPA/ATPDEA as temporary and inadequate; Peru actively seeks binding access | MINCETUR; Villarreal, 2006; Kat, 2023 | Hoop test against coercion-only explanation |

| <b>Mechanism step</b>                             | <b>Observable implications</b>                                                                                                          | <b>Principal sources</b>                                                             | <b>Inferential leverage</b>                            |
|---------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|--------------------------------------------------------|
| U.S. strategic framing of the Andean negotiation  | Official U.S. documents link the FTA to democracy, stability, and counternarcotics as well as trade                                     | USTR, 2003; Taft-Morales, 2009                                                       | Smoking-gun against purely economic explanation        |
| Deep agreement and domestic legal adaptation      | Treaty includes investment, IP, labor, environment, procurement, services, and dispute settlement; implementation requires legal reform | USTR, 2009; USTR, n.d.-a; Bolle & Villarreal, 2007; Canessa Montejo, 2020            | Hoop and supporting evidence for institutional lock-in |
| Post-ratification diplomatic and regional effects | Peru sustains close cooperation with Washington, accepts CAN fragmentation, and preserves the treaty across administrations             | St John, 2017; Ramírez Bullón, 2018; Fairlie, Collantes, & Castillo, 2021; Kat, 2023 | Cumulative straw-in-the-wind evidence                  |

*Note. Own elaboration.*

## **From ATPA to the PTPA**

The PTPA can only be understood within Peru's broader post-1990 strategy of international insertion. In the aftermath of the crisis and hyperinflation of the late 1980s, Peru underwent a profound shift toward market liberalization, trade openness, and the pursuit of external credibility. The Fujimori period consolidated that turn institutionally, but it also left a weakened democracy and a foreign-policy reputation marked by authoritarianism and corruption. When Alejandro Toledo assumed office in 2001, he inherited both an outward-oriented economic framework and the political imperative of restoring Peru's legitimacy abroad. His government consequently sought to combine democratic reconstruction, foreign investment, open markets, and closer relations with the United States, while also reinforcing Peru's multilateral engagement and Pacific projection (McClintock & Vallas, 2010; Rojas, 1997; St John, 2017).

ATPA and ATPDEA provided the immediate commercial backdrop to the treaty. Enacted in 1991 and later expanded through ATPDEA in 2002, these programs granted unilateral tariff preferences to Peru, Bolivia, Colombia, and Ecuador. Their rationale was explicitly tied to the U.S. counternarcotics agenda by encouraging legal export activity in the Andean region, Washington sought to promote alternatives to coca cultivation and illicit trade. For Peru,

these arrangements opened significant opportunities, particularly in textiles and agro-industrial exports, yet those gains rested on an inherently uncertain framework. Because the preferences depended on periodic renewal by the U.S. Congress, remained exposed to shifts in political priorities in Washington, and offered limited long-term predictability for investors and export sectors, they fell short of providing a durable basis for external economic strategy. Peruvian official sources later underscored precisely this point where unilateral preferences had expanded opportunities, but they did not amount to a sufficiently stable foundation for an export-led strategy (MINCETUR, n.d.-a, n.d.-b; USTR, 2009).

For Peru, the central issue was not whether access to the U.S. market mattered, but how that access could be stabilized and insulated from political uncertainty. That concern became more pressing in the early 2000s as the country sought deeper integration into global production networks while attracting investment under democratic rule. A reciprocal agreement appeared to offer an answer. Beyond safeguarding preferences vulnerable to expiration, it signaled more broadly that Peru was a reliable, rules-based destination for investment and a stable partner in international economic governance. In this sense, commercial interests and reputational considerations were not separate dynamics, but mutually reinforcing dimensions of the same strategy.



The U.S. side of the story was no less strategic. By the early 2000s, the broader FTAA project had begun to lose traction amid resistance from Brazil, Venezuela, and other governments skeptical of hemispheric integration under U.S. leadership. Washington responded by deepening a bilateral and subregional approach widely associated with competitive liberalization (Phillips, 2005; Quiliconi & Wise, 2009). In November 2003, U.S. Trade Representative Robert Zoellick formally notified Congress of the administration's intention to launch free-trade negotiations with Colombia, Peru, Ecuador, and Bolivia. Significantly, the initiative was framed in terms far broader than commercial liberalization alone. The notification linked the proposed Andean FTA to economic growth, investment, democratic governance, and the region's "special challenges," tying trade policy to a wider post-9/11 security and geopolitical agenda (USTR, 2003).

Timing was critical because the Andean region was undergoing a visible political realignment. Peru and Colombia moved toward closer cooperation with Washington, whereas Venezuela, Bolivia, and eventually Ecuador gravitated toward more skeptical, and at times explicitly anti-U.S., positions. In that context, trade negotiations unfolded in a regional environment where commercial policy and geopolitical ordering had become deeply intertwined. Peruvian leaders did not treat the treaty as a narrow instrument of market access. Rather, they approached it as part of a broader strategy of international repositioning through which Peru

reinforced its liberal trajectory while marking distance from competing regional projects (Kat, 2023; Ramírez Bullón, 2018; St John, 2017).

The sequence through which the treaty was negotiated and implemented itself reflects this layered political context. Negotiations formally commenced in 2004 after preliminary discussions the previous year. Peru and the United States completed bilateral negotiations in December 2005, signed the agreement in Washington in April 2006, adjusted the text in 2007 to incorporate the labor and environmental provisions associated with the U.S. congressional “May 10” framework, and eventually brought the treaty into effect on 1 February 2009. Throughout this process, the agreement was presented not merely as an instrument for trade liberalization, but as a far-reaching accord governing investment, services, government procurement, intellectual property, labor, environmental standards, transparency, and dispute settlement (MINCETUR, n.d.-c; USTR, n.d.-a; Villarreal, 2006).

The negotiation process itself reinforced Peru’s shift toward a differentiated form of regional insertion. Bargaining unfolded across multiple rounds in both the United States and Andean venues, addressing issues ranging from agriculture and rules of origin to services, intellectual property, dispute settlement, labor, and environmental standards. Although the initiative formally began as a joint negotiation with the

Andean countries, differences in bargaining pace and political viability soon became pronounced. Peru's decision to continue advancing as a bilateral negotiating partner once collective momentum weakened is significant in itself. It suggested that Lima prioritized securing an agreement with Washington over maintaining a common Andean negotiating front, a choice with immediate consequences for regional politics (Kat, 2023; Villarreal, 2006).

This institutional depth is central to the argument advanced here. Had the agreement been limited to tariff reduction alone, its implications for domestic institutions and foreign-policy orientation would likely have been far more constrained. Instead, the PTPA extended across a broad range of policy areas and required adjustments throughout multiple agencies of the Peruvian state. This helps explain why the agreement became both a focal point of domestic political contestation and a marker of Peru's preferred trajectory within globalization. It also clarifies why the treaty's political effects endured beyond the government that originally negotiated it.

Table 2 presents the critical chronology of the agreement and highlights the relevance of each episode to the causal argument.

**Table 2***Critical chronology of the Peru-U.S. treaty*

| <b>Date</b>  | <b>Event</b>                                               | <b>Relevance to the argument</b>                                                                                        |
|--------------|------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| 1991         | ATPA enacted                                               | Created unilateral, trade-for-counternarcotics preferences and an early pattern of dependence on U.S. renewal decisions |
| 2002         | ATPDEA enacted                                             | Expanded and renewed preferences, increasing the value (but also the political insecurity) of access to the U.S. market |
| 18 Nov. 2003 | USTR notifies Congress of Andean FTA negotiations          | Officially ties the proposed negotiation to trade, democracy, governance, and regional stability                        |
| May 2004     | First negotiating round                                    | Launches the formal bargaining process and opens the domestic coalition-building phase in Peru                          |
| 7 Dec. 2005  | Peru and the United States conclude bilateral negotiations | Confirms Peru's willingness to proceed even as broader Andean coordination weakens                                      |

| <b>Date</b>        | <b>Event</b>                                  | <b>Relevance to the argument</b>                                                                |
|--------------------|-----------------------------------------------|-------------------------------------------------------------------------------------------------|
| 12<br>Apr.<br>2006 | Treaty signed in<br>Washington                | Transforms negotiation into a<br>binding political commitment                                   |
| 28<br>Jun.<br>2006 | Peruvian Congress<br>approves treaty          | Signals broad elite support for the<br>agreement despite domestic<br>opposition                 |
| 2007               | Protocol of<br>amendment and U.S.<br>approval | Deepens labor and environmental<br>obligations and enables final<br>ratification in Washington  |
| 1 Feb.<br>2009     | PTPA enters into<br>force                     | Begins the implementation phase<br>and the observable foreign-policy<br>afterlife of the treaty |

*Note. Own elaboration.*

## **Process-Tracing the Agreement**

### *Mechanism 1 - Preference Insecurity and Coalition Formation in Peru*

The first stage of the mechanism lies in Peru's own search for permanence and credibility. As Kat (2023) shows,

Peru actively pursued participation in negotiations that had initially been conceived in broader Andean terms. This aligns with official Peruvian documentation, which consistently emphasized that unilateral preferences did not provide a sufficient foundation for long-term export policy and that a comprehensive agreement would offer a more stable setting for investment (MINCETUR, n.d.-a, n.d.-b). This point is more than incidental. It suggests that Peru was not merely responding to an initiative from Washington but advancing a strategy that connected trade policy to a broader project of economic insertion and diplomatic reliability.

That demand was translated into domestic coalition-building. MINCETUR's public account of the negotiations underscores the role of the *Sala Adjunta*, a consultative mechanism through which business associations, labor organizations, universities, regional governments, civil society groups, political parties, and congressional actors both received information and contributed proposals. The government also established the *Consejo Empresarial de Negociaciones Internacionales* to coordinate business participation (MINCETUR, n.d.-b). In two-level-games terms, these mechanisms expanded the domestic win-set surrounding the agreement while enabling negotiators to present their positions as expressions of national interests rather than merely executive preferences.

The coalition's support also rested on material interests. Textile producers, agro-exporters, and internationally oriented business associations had clear incentives to favor permanence over reliance on unilateral renewal cycles. A reciprocal agreement offered greater predictability for contracts, investment decisions, and long-term access to the U.S. market. In the Peruvian political context, those promises also carried reputational significance. In the aftermath of the Fujimori crisis, the treaty was presented as evidence that Peru was no longer a risky or erratic country, but a state capable of undertaking demanding, rule-based commitments with the hemisphere's leading power. The domestic coalition, in turn, was driven by the convergence of sectoral interests and national-image politics.

The coalition was neither homogeneous nor uncontested. Issues involving agriculture, labor standards, investor protections, public procurement, and especially intellectual-property provisions generated important domestic resistance. Von Braun's (2012) study of the IP negotiations shows how public-health concerns and sectoral actors challenged the scope of U.S. demands regarding data exclusivity, patents, and pharmaceutical regulation. That conflict is analytically revealing because it exposes the character of the bargain itself. A government willing to accept politically costly provisions on medicines, copyrights, and enforcement is doing more than maximizing short-term

export gains; it is making a broader strategic choice about how to embed the national economy and which foreign partnerships to privilege.

Ramírez Bullón (2018) introduces a critical foreign-policy layer to this coalition logic by showing that the treaty was constructed domestically not only as an economic imperative, but also as part of a broader effort to legitimize free trade within Peruvian statecraft. In this framing, the agreement came to embody seriousness, modernization, and international credibility. The central actors were not simply exporters seeking sectoral gains, but technocrats and political elites who viewed trade policy as intertwined with foreign policy and treated the treaty as a vehicle for positioning Peru within the international order. This interpretive lens matters because it helps explain both why the agreement endured despite intense contestation and why its political symbolism reached well beyond its tariff provisions.

This initial stage of the mechanism casts serious doubt on purely coercive accounts. Peruvian leaders and trade technocrats did not approach the treaty simply as an external imposition, but as an instrument for securing material gains, projecting reliability to investors, and embedding Peru within a rules-based relationship with the hemisphere's dominant power. In Schweller's terms, the underlying logic was closer to bandwagoning for profit than passive compliance: Peru



moved toward alignment in pursuit of expected benefits, rather than because it lacked meaningful alternatives.

### *Mechanism 2 - U.S. Competitive Liberalization and Andean Strategy*

The second stage centers on the way the United States strategically defined the negotiation itself. Zoellick's 2003 letter to Congress is particularly significant because it explicitly presented an Andean FTA as supporting democracy, accountable governance, the rule of law, labor rights, anti-corruption efforts, and a regional response to narcotrafficking, while also advancing competitive liberalization and serving broader hemispheric objectives (USTR, 2003). Because this evidence appears in the formal notification that initiated the negotiations, it carries unusual inferential weight. It is difficult to square such evidence with accounts that portray U.S. motivations as exclusively economic.

This strategic framing is consistent with wider scholarship on U.S. bilateralism in the hemisphere. Shadlen (2008) argues that trade agreements in the Americas reflected both power asymmetries and the strategic deployment of bilateralism after the FTAA lost momentum. Phillips (2005) and Quiliconi and Wise (2009) similarly show that bilateral trade governance enabled Washington to proceed with willing partners when broader regional

initiatives stalled. Kat (2023) applies this logic to Peru, showing that competitive liberalization also functioned to contain postneoliberal regionalism. Peru proved particularly valuable because it combined a cooperative executive, reliance on U.S. preferences, and a political discourse that portrayed liberalization as modernization rather than subordination.

Counternarcotics considerations further reinforced this convergence. ATPA and ATPDEA had long-tied trade preferences to anti-drug objectives, granting market access as part of a strategy aimed at promoting legal production alternatives in the Andean region (USTR, 2009). The transition from unilateral preferences to a reciprocal and permanent agreement did not displace that logic; rather, it incorporated it into a wider institutional framework. In this sense, the treaty carried forward the political rationale of the preference regime while elevating the level of commitment on both sides. The same negotiation that ensured more secure market access also entailed cooperation on governance, legal reform, and regional stability (Taft-Morales, 2009).

The regional setting amplified the significance of the agreement. Bolivia and Ecuador moved progressively away from the U.S. trade agenda, while Venezuela turned opposition to such agreements into a central element of its regional discourse. In this context, Peru and Colombia emerged as the most viable Andean partners for Washington's

preferred model of hemispheric economic governance. The importance of the PTPA lay not in generating ideological polarization on its own, but in translating an existing divide into a concrete institutional commitment. This helps explain why the agreement resonated well beyond the domain of tariff liberalization.

At this stage, the mechanism does not reflect hidden geopolitical motives masked by an economic façade. Rather, it points to openly intertwined objectives. The treaty held particular value for Washington because it brought together market access, rule-making, counternarcotics, and political influence in a region where competing integration projects were emerging. This constitutes the article's strongest smoking-gun evidence against a purely economic interpretation.

### *Mechanism 3 - Deep Agreement and Institutional Lock-In*

The third stage focuses on the domestic requirements generated by the treaty. The PTPA constituted a deep agreement. Official Peruvian sources describe it as encompassing investment, intellectual property, labor, environment, services, public procurement, competition policy, transparency, customs, and dispute settlement alongside trade in goods (MINCETUR n.d.-a, n.d.-c). This breadth is important because deep agreements do more than lower tariffs: they reallocate authority across ministries,

empower some domestic actors over others, and increase the costs of policy reversal. Once this is acknowledged, the treaty's foreign-policy implications become much clearer.

The implementation record reinforces this claim. USTR's 2009 report indicates that Peru modified key elements of its legal framework to comply with obligations related to intellectual-property protection and enforcement. The 2007 amendment process, required in the United States following congressional negotiations over labor and environmental provisions, also strengthened the agreement's social-policy chapters and made it the first U.S. FTA in force to adopt that template (Bolle & Villarreal, 2007; USTR, n.d.-a). The details of each individual decree are less important here than the broader point. Ratification of the agreement required Peru to translate external commitments into domestic law across multiple policy domains. In process-tracing terms, this marks the stage at which an external commitment becomes internalized within national institutions.

This process of internalization produced clear distributive and diplomatic effects. Once the agreement was incorporated into domestic law and implemented by state agencies, it generated constituencies favoring continuity, where export sectors benefiting from stable market access, investors valuing legal predictability, ministries assuming new responsibilities, and political elites linking national

credibility to compliance. In other words, the treaty reshaped the domestic dimension of the two-level game. Subsequent governments did not operate on a blank slate; they confronted an agreement whose reversal would entail economic, legal, and diplomatic costs.

At the same time, the limits of the treaty's social effects must also be acknowledged. Canessa Montejo (2020) argues that the labor chapter did not automatically or comprehensively transform labor conditions in Peru. The provision was significant, but its impact was mediated by preexisting labor-market structures and by the asymmetrical character of the agreement. This serves as a useful reminder that institutional lock-in should not be overstated. The treaty both constrained and enabled it generated external credibility and commercial opportunities, while simultaneously reducing the scope for alternative regulatory and developmental strategies.

Ratification politics further reinforced this lock-in at a symbolic level. The Peruvian Congress approved the agreement by a wide margin, and Alan García publicly defended it against nationalist and leftist critics, framing the treaty as part of a broader choice about Peru's development model and regional role (Kat, 2023; St John, 2017; USTR, 2009). The PTPA thus functioned both as a policy instrument and as a marker of identity, where an external commitment that helped define the type of regional actor Peru sought to

become. The continuity explanation retains some plausibility (García did not originate Peru's liberal turn) but the evidence indicates that the treaty went beyond merely reflecting a preexisting orientation. It embedded that orientation in law, institutions, and distributive coalitions, thereby granting it greater durability than presidential rhetoric alone could sustain.

## **Foreign-Policy Effects of the Agreement**

### *Counternarcotics and the Security Agenda*

The first foreign-policy effect is more accurately characterized as consolidation rather than innovation. Peru and the United States were already cooperating on counternarcotics before the PTPA, and the ATPA/ATPDEA framework had itself been grounded in anti-drug policy objectives. The treaty's contribution was to solidify that relationship, making it more durable and more closely integrated into a wider governance agenda. U.S. official discourse framed trade, democracy, and narcotics control as mutually reinforcing, while Peruvian authorities presented the agreement as part of a stable and modern international strategy (Taft-Morales, 2009; USTR, 2003, 2009).

This point is important because it highlights the convergence of economic and security agendas in the Andean context. Access to the U.S. market functioned not only as a

commercial benefit, but also as a marker of political alignment and institutional trust. Although the treaty did not include explicit security provisions, it deepened the bilateral relationship by embedding commercial integration within a broader framework of cooperation. For this reason, the PTPA is best interpreted as a foreign-policy instrument, despite being formally structured as a trade agreement.

### *Regional Integration and the Fragmentation of the Andean Space*

The second effect pertains to regional integration. The PTPA further eroded expectations that the Andean Community could act or develop as a unified negotiating bloc vis-à-vis the United States. While negotiations initially took place in an Andean framework, Peru eventually advanced on a bilateral track as collective efforts faltered. The political implications of this fragmentation became especially apparent when Venezuela withdrew from the Andean Community, explicitly in response to the FTAs pursued by Peru and Colombia with Washington (St John, 2017; Villarreal, 2006). Although the treaty did not on its own precipitate the crisis of Andean integration, it served as a catalytic turning point within an increasingly polarized ideological landscape.

The Pacific dimension of Peru's foreign policy became more prominent following the treaty. The PTPA contributed

to normalizing the view that Peru's most viable international trajectory lay not in preserving the cohesion of a politically fragmented Andean bloc, but in engaging in open regionalism linked to global markets across the Pacific basin. This did not imply abandoning South America, but it did entail a reordering of priorities. The subsequent emergence of the Pacific Alliance cannot be attributed solely to the PTPA, yet the agreement clearly helped lay the institutional and discursive foundations for a foreign policy that prioritized commercial openness, regulatory alignment, and extra-regional connections over ideological solidarity.

Recent trade scholarship introduces an important qualification. Fairlie, Collantes, and Castillo (2021) show that intra- and extra-regional agreements did not significantly shape trade flows within the Andean Community. This does not undermine the present argument; rather, it sharpens it. The significance of the PTPA did not lie primarily in mechanically redirecting trade away from the Andean market. Instead, its importance was political and institutional: it weakened the credibility of a common Andean negotiating position and normalized differentiated strategies of international insertion. In this sense, the treaty contributed to repositioning Peru away from a contested Andean regionalism and toward a more Pacific-oriented and liberal trajectory.



This repositioning becomes evident in subsequent foreign-policy developments. Ramírez Bullón (2018) persuasively shows that the PTPA was part of a broader process through which free trade was legitimized within Peruvian diplomacy, ultimately helping to set the stage for the Pacific Alliance. The treaty did not mechanically generate the Alliance, but it reinforced a state discourse in which integration was conceived as outward-oriented, Pacific-focused, and anchored in extra-regional agreements rather than in protective or inward-looking regionalism. The political afterlife of the PTPA therefore extended beyond the bilateral relationship itself.

### *Peru and the Bolivarian Challenge*

A third consequence for foreign policy involved distancing the state, both strategically and symbolically, from the Bolivarian axis. As St John (2017) demonstrates, the Toledo and García administrations increasingly prioritized pragmatic global integration, democratic norms, and open markets, standing in stark contrast to the alternative regional paradigms championed by Venezuela, Bolivia, and occasionally Ecuador. Within this context, the PTPA operated as an external commitment device. The agreement communicated that the country's favored mode of international insertion relied upon alignment with Washington, adherence to liberalized commercial regimes, and membership within a bloc of market-oriented nations.

This demarcation was rendered acutely overt through García's political discourse, particularly his condemnation of what he termed "Andean fundamentalism."

However, the empirical record simultaneously reveals boundaries holding substantial theoretical significance. Lima avoided totally rupturing diplomatic channels with leftist hemispheric actors, declining to transform this accord into a singular partnership alongside the United States. Foreign policy maintained pragmatic orientations; subsequent administrations recalibrated discursive framing whilst keeping core commitments intact. The resilience characterizing this policy foundation emerged profoundly during Ollanta Humala's administration. Notwithstanding prior hostility toward this pact alongside discursive alignments with specific leftist factions, his presidency eventually functioned inside the country's institutionalized commercial architecture, actively stepping back from the highly antagonistic regional posture favored by Chávez (Kat, 2023). Consequently, the PTPA constrained feasible policy trajectories without entirely eradicating sovereign space for political maneuver. Such dynamics perfectly embody what this study identifies as selective, rather than unconditional, bandwagoning.

## **Economic Outcomes, Distribution, and Structural Limits**

While preceding analytical sections demonstrate the PTPA's undeniable political significance, evaluating its underlying economic outcomes remains equally critical, given that commercial impacts directly consolidated the agreement's institutional legitimacy while simultaneously bolstering those internal coalitions advocating for policy continuity. State statistical records confirm the pact yielded highly observable material dividends. Over the accord's initial fifteen years, MINCETUR reports that total shipments bound for the United States exceeded the US\$101 billion threshold. Outbound trade hovering around US\$4.7 billion in 2009 successfully surpassed US\$9 billion by 2023; this upward trajectory cemented the domestic conviction that permanent structural integration into American markets constituted an invaluable sovereign asset (MINCETUR, 2024). Furthermore, official American governmental metrics confirm Lima's enduring status as a vital bilateral counterpart, with combined goods and services exchanges reaching an estimated US\$26.7 billion throughout 2024 (USTR, n.d.-b).

Beyond aggregate macroeconomic gains, these material dividends were concurrently matched by significant structural broadening across commercial sectors. MINCETUR's sixteenth-year evaluation documents that

1,806 unprecedented tariff subheadings successfully reached the United States throughout the accord's lifespan, generating overall revenues approaching US\$9.94 billion, overwhelmingly clustered outside conventional mining or energy industries. The official spotlighting of textiles, chemicals, metal-mechanical goods, and related nontraditional outflows holds considerable analytical weight because it implies the institutional framework transcended mere enlargement of the nation's existing mineral basket (MINCETUR, 2025). While this scarcely implies that Lima relinquished its structural reliance upon primary commodities, it strictly confirms the agreement widened viable pathways within industries highly functional for the government's political narrative of modernization and export diversification.

Academic literature generally corroborates the premise that this accord transformed structural trade configurations, even if the scale and developmental caliber of such shifts remain fiercely contested. Utilizing highly disaggregated data, Arrieta Padilla's gravity-model analysis examines trade creation and diversion effects within the framework, ultimately determining that the PTPA shifted bilateral commercial dynamics through mechanisms irreducible to mere macroeconomic trends (Arrieta Padilla, 2017). Such rigorous empirical substantiation proves analytically crucial by reinforcing the assertion that the pact possessed genuine commercial efficacy. Simultaneously,

institutional proponents of the agreement frequently leaped prematurely from raw trade expansion toward developmental triumphs, proceeding as though export growth inherently resolved broader debates concerning the accord's overarching domestic worth.

A more skeptical analytical perspective highlights enduring structural constraints. Alarco Tosoni (2017) contends that free trade frameworks across Chile, Mexico, and Peru failed to independently generate definitive ruptures within their GDP growth trajectories. Furthermore, both Peru and Chile persisted in exporting numerous goods toward the United States characterized by minimal processing alongside restricted technological sophistication. This constitutes a vital caveat. Nations might successfully expand outbound trade while fundamentally failing to deeply reorganize their internal productive apparatus. Consequently, although the PTPA fostered international insertion, sovereign credibility, and marginal diversification, it never intrinsically catalyzed advanced technological upgrading, sophisticated industrial manufacturing, or entirely novel developmental paradigms.

A comparative framework further clarifies this dynamic. Alarco Tosoni (2017) juxtaposes the Andean nations against Mexico, demonstrating that whereas the latter achieved deep manufacturing integration with Washington, both Peru and Chile stayed disproportionately tethered to primary commodities and products featuring minimal

transformation. Such distinctions hold theoretical weight, considering domestic proponents frequently justified the PTPA by assuming unfettered American market entry would intrinsically propel the nation into sophisticated global value chains. Available empirical records indicate far humbler realities. While Lima successfully secured baseline access and peripheral export diversification, the state ultimately failed to duplicate the profound industrial articulation characterizing highly manufacturing-intensive North American commercial integration.

Consequently, evaluating this export basket's structural quality remains paramount. Merely observing absolute outbound trade growth or acknowledging how nontraditional merchandise proliferated proves analytically insufficient. The more rigorous inquiry asks whether such advances catalyzed elevated technological sophistication, profound domestic linkages, alongside mitigated reliance upon highly volatile macroeconomic sectors. Regarding these specific parameters, empirical results remain decidedly ambivalent. This accord successfully anchored market entry, streamlined strategic foresight, and diversified commercial portfolios directed toward Washington. However, it could never replace comprehensive industrial policy, robust innovation policy, or structurally inclusive labor-market reforms. Therefore, this study emphasizes the pact's profound foreign-policy magnitude, because the PTPA possessed

immense strategic weight although its underlying tangible economic victories remained strictly partial.

Such ambivalent records concurrently illuminate the fundamental asymmetry of gains. Washington secured a rule-intensive framework safeguarding capital investment, consolidating regulatory disciplines, and projecting structural influence via legal mechanisms. Conversely, Lima acquired permanence, predictability, and lucrative commercial openings, albeit by accepting constraints upon policy discretion carrying undeniable systemic costs (Manger, 2009; Shadlen, 2005). Consequently, the treaty's economic afterlife structurally complements, rather than contradicts, this core argument. Commercial gains proved robust enough to politically anchor the agreement yet lacked sufficient transformative depth to render macroeconomic success the entire explanatory narrative.

Table 3 summarizes the main economic effects and limitations most relevant to the argument.

**Table 3***Economic outcomes most relevant to the argument*

| <b>Indicator</b>                 | <b>Official or scholarly evidence</b>                                                                                        | <b>Interpretation for the argument</b>                                                                         |
|----------------------------------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| Cumulative export growth         | MINCETUR reports that exports to the United States surpassed US\$101 billion in the first fifteen years of the treaty        | The agreement delivered visible commercial gains and strengthened domestic support for continuity              |
| Diversification by product lines | MINCETUR's sixteenth-year assessment identifies 1,806 new tariff subheadings and strong non-mining, non-energy participation | The treaty widened opportunities beyond traditional commodity exports, even if structural dependence persisted |
| Ongoing bilateral trade weight   | USTR reports estimated U.S.-Peru goods and services                                                                          | The bilateral relationship remained economically central well after entry into force                           |



| Indicator         | Official or scholarly evidence             | Interpretation for the argument                                                                                       |
|-------------------|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
|                   | trade of US\$26.7 billion in 2024          |                                                                                                                       |
| Structural limits | Alarco Tosoni, 2017; Canessa Montejo, 2020 | Export growth and legal lock-in did not automatically generate technological upgrading or broad social transformation |

*Note. Own elaboration.*

## Discussion

Process-tracing evidence inherently substantiates a multifaceted analytical interpretation. This PTPA functioned beyond mere export agreements, yet it simultaneously resisted reduction toward straightforward narratives of American hegemonic imposition. Peru proactively pursued this accord anticipating lucrative market access, heightened investor confidence, alongside elevated international claims regarding sovereign reliability and institutional modernity. Washington advanced the pact foreseeing reciprocal commercial dividends combined with strategic hemispheric influence amidst profound regional ideological contestation.

Ultimately, the agreement's deep institutional design effectively translated such mutual preferences into a durable structure of commitment.

Such structural dynamics are optimally conceptualized as selective bandwagoning. Rather than balancing against the dominant power, Peru deliberately integrated into that hegemon's favored regional architecture, securing tangible material and political rewards. Nevertheless, this specific bandwagoning categorization warrants analytical caution. The Andean nation remained far from a strictly passive client. Domestic political actors actively negotiated, fiercely contested, and strategically appropriated the resulting framework. Consequently, these results ultimately constitute a negotiated form of asymmetric alignment, an approach where a weaker state seeks prestige and advantages by embedding itself within a rule-based relationship alongside a stronger partner.

This empirical case concurrently illuminates Peru's overarching trajectory throughout globalization. This pact constituted substantially more than merely ratifying another routine commercial arrangement. It consolidated an expansive sovereign project wherein free trade, diplomatic pragmatism, and Pacific-facing integration functioned as mutually reinforcing pillars underpinning national strategy. Ramírez Bullón's (2018) research proves highly instructive regarding this dynamic; rather than simply accompanying the

ideological legitimization of free trade within Peruvian foreign policy, the PTPA actively helped institutionalize it. The accord provided legal density alongside symbolic weight to an insertion model subsequently structuring Peru's hemispheric selections and public self-presentation as an open, highly reliable, investor-friendly state.

Competing hypotheses can subsequently be evaluated with significantly greater precision. Strictly macroeconomic paradigms prove demonstrably inadequate, given that decisive American governmental documentation explicitly linked the bargaining process to democratic stability, governance, and counternarcotics. The continuity narrative remains partially valid though structurally deficient; while a cooperative, market-friendly orientation predated this pact, the agreement ultimately entrenched and institutionalized it. Ideological frameworks encapsulate fragments of García's discourse, particularly regarding Chávez; however, they fail to elucidate the accord's survival across administrations featuring divergent discursive approaches. Unifying these diverse empirical findings is the causal mechanism elaborated previously: preference insecurity generated demand, Washington's strategy supplied opportunity, and deep institutionalization guaranteed durability.

More expansively, this empirical observation implies that the foreign-policy significance characterizing trade agreements rests upon three foundational conditions:

asymmetry, depth, and regional contestation. Whenever a weaker state formalizes a deep agreement alongside a dominant power amidst competing regional projects, commercial policy effectively transforms into an instrument of strategic alignment. Absent these specific prerequisites, comparable institutional frameworks might exert diminished geopolitical effect. Consequently, the Peruvian case proves profoundly theoretically useful far beyond the Andes. It demonstrates that the politics of globalization remain inadequately understood whenever bilateral treaties are exclusively evaluated through narrow tariff schedules and raw trade volumes.

These findings possess broader ramifications regarding how scholars frequently conceptualize Latin American international political economy. Frameworks artificially decoupling commercial strategy from external relations persistently overlook how these domains operate as mutually constitutive arenas across numerous administrations. Regarding Peru, resolving to anchor domestic markets deeper within U.S.-centered regulatory architectures simultaneously represented a profound recalibration of diplomatic identity, regional coalition-building, and tolerable parameters of sovereign dependence. Consequently, this accord demands interpretation not as a peripheral foreign policy appendage, but strictly as a primary technological instrument of statecraft. Such analytical takeaways potentially extend toward comparable scenarios

where deep agreements effectively replace formal alliance politics, operating instead as vital strategic mechanisms for repositioning states amidst contested regional orders.

This empirical record complicates reductionist debates contrasting autonomy and dependence across Latin America. Peru resisted total absorption into Washington's overarching agenda, yet it similarly failed to maintain undisturbed policy autonomy following this accord. Rather, the state pursued autonomy through association; because it actively chased sovereign maneuverability, international credibility, alongside economic opportunity by anchoring itself deeply toward a dominant partner's legal and regional order. Consequently, the PTPA proves fundamentally crucial as an instructive foreign-policy case. It powerfully reveals how nations can adopt strategies of insertion remaining simultaneously agentic and constraining.

Ultimately, this research yields vital methodological implications. Prevailing scholarly tendencies frequently catalog the agreement's broader environment (the ATPA, the Andean crisis, U.S. strategy, alongside export performance) without rigorously demonstrating exactly how these isolated variables interact causally. Process tracing structurally demands a profoundly disciplined narrative. Once this sequence is methodically mapped, the pact entirely ceases resembling any generic by-product of globalization. Rather, it emerges as a historically specific political instrument whereby

economic governance, regional order, and foreign-policy alignment become mutually reinforcing.

## **Conclusion**

This manuscript aimed to elucidate how the Peru-United States Trade Promotion Agreement reconfigured Peruvian foreign policy, employing process tracing instead of descriptive narratives. Empirical findings substantiate one definitive conclusion. The PTPA transcended operating purely as an economic mechanism driving trade. It constituted a politically consequential institution simultaneously reflecting and actively reinforcing Peru's structural alignment with the United States across the Andean regional context.

Three causal mechanisms elucidate this specific outcome. Initially, the inherent insecurity of ATPA/ATPDEA preferences generated powerful internal motivations across Peru toward securing perpetual market access alongside investment credibility, thereby mobilizing domestic coalitions advocating comprehensive treaty ratification. Furthermore, the United States structured and advanced this bargaining process within broader hemispheric designs prioritizing competitive liberalization, democracy promotion, and rigorous counternarcotics governance. Finally, the accord's depth necessitated systemic regulatory transformations that structurally locked in bilateral relations, simultaneously

fortifying those actors deeply invested in its institutional continuity.

This manuscript's wider implication contends that commercial pacts demand evaluation beyond mere export volumes or strictly macroeconomic welfare effects. Across asymmetrical regional settings, deep trade agreements effectively operate as instruments of geopolitical ordering. Beyond allocating material gains, these frameworks allocate systemic influence, reshape coalition politics, whilst restricting foreign-policy alternatives. The Peru-U.S. case illustrates such dynamics impeccably, given that the accord materialized amidst intersecting market insecurity, profound regional polarization, alongside broader U.S. hemispheric strategy.

Simultaneously, available macroeconomic evidence strictly cautions against embracing simplistic triumphalist narratives. This pact stimulated commercial expansion, facilitated peripheral export diversification, while concurrently bolstering internal validation regarding Peru's prevailing outward-oriented state strategy. However, it failed to independently catalyze definitive structural transformation, advanced technological upgrading, or comprehensive social inclusion. This ambivalent legacy remains foundational to the manuscript's claim. Ultimately, the PTPA persisted not by definitively resolving every

Peruvian developmental dilemma, but through successfully anchoring a durable model of international insertion.

Subsequent scholarship could corroborate and elevate this thesis through integrating diplomatic repositories, legislative committee transcripts, alongside direct dialogues involving Peruvian and American negotiators. Cross-national analyses involving Colombia, alongside alternative scenarios wherein comparable commercial frameworks yielded diminished institutional impacts, would similarly prove highly instructive. Notwithstanding reliance upon publicly accessible documentation, this causal mechanism remains sufficiently transparent to sustain considerably bolder deductions than prevailing narratives concede, that the PTPA demands interpretation operating beyond mere commercial treaties, functioning profoundly as an instrument of strategic alignment successfully repositioning Peru across the broader political economy of the Americas.

## References

- Alarco Tosoni, G. A. (2017). Tratados de libre comercio, crecimiento y producto potencial en Chile, México y Perú. *Economía UNAM*, 14(42), 24–46.  
<https://doi.org/10.1016/j.eunam.2017.09.001>
- Arrieta Padilla, G. L. (2017). *Un análisis gravitacional de la creación y desviación comercial en el marco del Tratado de Libre Comercio entre Perú y Estados*



- Unidos* (Documento de Trabajo 2017-010). Banco Central de Reserva del Perú.
- Beach, D., & Pedersen, R. B. (2013). *Process-tracing methods: Foundations and guidelines*. University of Michigan Press.
- Bennett, A., & Checkel, J. T. (Eds.). (2015). *Process tracing: From metaphor to analytic tool*. Cambridge University Press.
- Bolle, M. J., & Villarreal, M. A. (2007). *Peru Trade Promotion Agreement: Labor issues* (CRS Report RS22521). Congressional Research Service.
- Büthe, T., & Milner, H. V. (2008). The politics of foreign direct investment into developing countries: Increasing FDI through international trade agreements? *American Journal of Political Science*, 52(4), 741–762. <https://doi.org/10.1111/j.1540-5907.2008.00340.x>
- Canessa Montejo, M. F. (2020). *El impacto del capítulo laboral del tratado de libre comercio entre Perú y Estados Unidos en el mundo del trabajo peruano*. Retrieved April 6, 2026, from <http://www.relatos.org/documentos/GLOB.Canessa.pdf>
- Cantuarias Salaverry, F., & Salvador Raygada, P. (2008). *TLC Perú-Estados Unidos: Contenido y aplicación*. Universidad Peruana de Ciencias Aplicadas.
- Collier, D. (2011). Understanding process tracing. *PS: Political Science and Politics*, 44(4), 823–830.

- Fairlie, A., Collantes, E., & Castillo, L. (2021). El rol de los acuerdos intra y extrarregionales en los flujos comerciales: El caso de la Comunidad Andina de Naciones. *Problemas del Desarrollo*, 52(204), 165–188.  
<https://doi.org/10.22201/IIEC.20078951E.2021.204.69603>
- George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. MIT Press.
- Kat, Q. B. (2023). Competitive liberalization, postneoliberalism, and hegemony: The case of the US-Peru Free Trade Agreement. *Latin American Politics and Society*, 65(1), 126–150.  
<https://doi.org/10.1017/lap.2022.33>
- Manger, M. S. (2009). *Investing in protection: The politics of preferential trade agreements between North and South*. Cambridge University Press.
- McClintock, C., & Vallas, F. (2010). The United States and Peru in the 2000s. In J. I. Domínguez & R. Fernández de Castro (Eds.), *Contemporary U.S.-Latin American relations: Cooperation or conflict in the 21st century* (pp. 197–217). Routledge.
- Ministerio de Comercio Exterior y Turismo. (n.d.-a). *Acuerdo de Promoción Comercial Perú-EE.UU.* Acuerdos Comerciales del Perú. Retrieved April 6, 2026, from [https://www.acuerdoscomerciales.gob.pe/En\\_Vigencia/EEUU/inicio.html](https://www.acuerdoscomerciales.gob.pe/En_Vigencia/EEUU/inicio.html)

- Ministerio de Comercio Exterior y Turismo. (n.d.-b). *Preguntas frecuentes: TLC Perú - EEUU*. Acuerdos Comerciales del Perú. Retrieved April 6, 2026, from [https://www.acuerdoscomerciales.gob.pe/En\\_Vigencia/EEUU/Preguntas\\_Frecuentes.html](https://www.acuerdoscomerciales.gob.pe/En_Vigencia/EEUU/Preguntas_Frecuentes.html)
- Ministerio de Comercio Exterior y Turismo. (n.d.-c). *Textos del acuerdo: TLC Perú - EEUU*. Acuerdos Comerciales del Perú. Retrieved April 6, 2026, from [https://www.acuerdoscomerciales.gob.pe/En\\_Vigencia/EEUU/Textos\\_Acuerdo.html](https://www.acuerdoscomerciales.gob.pe/En_Vigencia/EEUU/Textos_Acuerdo.html)
- Ministerio de Comercio Exterior y Turismo. (2024, August 2). *Tras 15 años del TLC, exportaciones peruanas a Estados Unidos superan los US\$ 101 mil millones* [Boletín]. <https://www.gob.pe/institucion/mincetur/noticias/998527-tras-15-anos-del-tlc-exportaciones-peruanas-a-estados-unidos-superan-los-us-101-mil-millones>
- Ministerio de Comercio Exterior y Turismo. (2025). *EEUU: 16° año de vigencia del TLC*. Dirección de Estudios Económicos, Ministerio de Comercio Exterior y Turismo.
- Phillips, N. (2005). U.S. power and the politics of economic governance in the Americas. *Latin American Politics and Society*, 47(4), 1–25.
- Putnam, R. D. (1988). Diplomacy and domestic politics: The logic of two-level games. *International Organization*, 42(3), 427–460.

- Quiliconi, C., & Wise, C. (2009). The US as a bilateral player: The impetus for asymmetric free trade agreements. In M. Solís, B. Stallings, & S. N. Katada (Eds.), *Competitive regionalism: FTA diffusion in the Pacific Rim* (pp. 97–117). Palgrave Macmillan.
- Ramírez Bullón, J. E. (2018). *La legitimación del libre comercio en la política exterior peruana: Del TLC con Estados Unidos al surgimiento de la Alianza del Pacífico* [Bachelor's thesis, Pontificia Universidad Católica del Perú].
- Rojas, J. (1997). La política comercial peruana reciente. *Economía*, 20(39–40), 207–248.
- Schweller, R. L. (1994). Bandwagoning for profit: Bringing the revisionist state back in. *International Security*, 19(1), 72–107. <https://doi.org/10.2307/2539149>
- Shadlen, K. C. (2005). Exchanging development for market access? Deep integration and industrial policy under multilateral and regional-bilateral trade agreements. *Review of International Political Economy*, 12(5), 750–775. <https://doi.org/10.1080/09692290500339685>
- Shadlen, K. C. (2008). Globalisation, power and integration: The political economy of regional and bilateral trade agreements in the Americas. *Journal of Development Studies*, 44(1), 1–20.
- St John, R. B. (2017). La política exterior peruana en el nuevo milenio: Continuidad y cambio. *Revista del Instituto*

Riva-Agüero, 2(2), 65–119.

<https://doi.org/10.18800/revistaira.201702.002>

Taft-Morales, M. (2009). *Peru: Current conditions and U.S. relations* (CRS Report R40716). Congressional Research Service.

United States Trade Representative. (n.d.-a). *Peru Trade Promotion Agreement*. Retrieved April 6, 2026, from <https://ustr.gov/trade-agreements/free-trade-agreements/peru-tpa>

United States Trade Representative. (n.d.-b). *Peru*. Countries and Regions. Retrieved April 6, 2026, from <https://ustr.gov/countries-regions/americas/peru>

United States Trade Representative. (2003, November 18). *Andean Free Trade Agreement* [Letter to Congress]. [https://ustr.gov/archive/Document\\_Library/Letters\\_to\\_Congress/2003/Andean Free Trade Agreement.html](https://ustr.gov/archive/Document_Library/Letters_to_Congress/2003/Andean_Free_Trade_Agreement.html)

United States Trade Representative. (2009). *Fourth report to the Congress on the operation of the Andean Trade Preference Act as amended*. <https://ustr.gov/sites/default/files/USTR%202009%20ATPA%20Report%20Final.pdf>

U.S. International Trade Commission. (2006). *U.S.-Peru Trade Promotion Agreement: Potential economywide and selected sectoral effects* (Publication 3855). <https://www.usitc.gov/publications/332/pub3855.pdf>

- Villarreal, M. A. (2006). *Andean-U.S. free-trade agreement negotiations* (CRS Report RL32770). Congressional Research Service.
- von Braun, J. (2012). *The domestic politics of negotiating international trade: Intellectual property rights in US-Colombia and US-Peru free trade agreements*. Routledge. <https://doi.org/10.4324/9780203156360>
- Walt, S. M. (1987). *The origins of alliances*. Cornell University Press.

# Epistemología Social



# Presentación de la sección “Epistemología social”

**Felipe Álvarez<sup>81</sup>**

Esta sección del libro se enfoca en otorgar distintas perspectivas contemporáneas en torno a la difusión de conocimiento y los problemas actuales que esto suscita. El artículo inaugural, “Responsabilidades éticas en la era de la posverdad” por Rodolfo Aldea, pone en la palestra el desafío estructural que suscita la posverdad en cuanto a las responsabilidades morales vinculadas al conocimiento colectivo. En este sentido, Aldea plantea que la cohesión social se pone en peligro si se amplifica la polarización de las opiniones y la desconfianza hacia la ciencia, por lo que es necesario desarrollar un enfoque interdisciplinario para

---

<sup>81</sup> Editor responsable del eje Epistemología social. Candidato a Doctor en Filosofía, Universidad de Chile. Investigador asociado, Fundación CIEPE — Centro Interdisciplinario en Ética, Política y Economía.  
<https://orcid.org/0000-0002-7153-2851>



abordar el fenómeno de la posverdad y sus efectos en la era digital.

El segundo artículo, “El conocimiento como bien común: una reflexión ética sobre su difusión en la era de la inteligencia artificial” por Eduardo León, sigue una línea similar; pero enfocado en la inteligencia artificial y el modo en que esta, dependiendo de cómo sea diseñada, puede suscitar la homogeneización cognitiva (entre otros peligros epistémicos), lo cual terminaría suscitando la concentración del poder en quienes tengan el control del algoritmo. Siendo esto así, León busca defender que el conocimiento como bien común no solo implicaría el acceso a este (el cual puede ser garantizado por el uso de las inteligencias artificiales junto con otros medios), sino también suscita que los agentes epistémicos sean partícipes de las condiciones que legitiman la validación y difusión del conocimiento. Como se podrá entrever, sin un marco de referencia ético para ello, el uso de la inteligencia artificial generativa puede derivar en prácticas que inhiben su potencial como herramienta para la democratización del conocimiento. Se buscará, pues, precisamente lo contrario.

El tercer capítulo, “Ética y conocimiento: el problema de la discriminación de los testimonios y los deepfakes” aborda en uno de los posibles usos nocivos de la inteligencia artificial generativa: los *deepfakes* y su posibilidad de ser usados para discriminar testimonios como si fueran una

evidencia en video. Según Álvarez, los *deepfakes* pueden mermar la posibilidad de conocer a través de la palabra de otros si se presentan *como si fueran evidencia* para perpetuar injusticias epistémicas. Al mismo tiempo, se pone el énfasis en las complicaciones éticas del uso de la imagen de una persona para generar material audiovisual ilegítimo que termine afectando su credibilidad.

En el cuarto capítulo, el problema de los *deepfakes* se profundiza al señalar que lo posible y lo real tienden a mimetizarse gracias al uso de esta tecnología, mermando la confianza social en los agentes políticos. Así, en “Deepfake y regímenes de visibilidad: Lula, Bolsonaro y el gesto inexistente” por Adriano Cruz, se propone que, si los *deepfakes* generan escenarios políticos ficticios verosímiles, pueden generar reacciones afectivas intensas en las personas, desestabilizando el imaginario político y mostrando su potencial para operar como dispositivos mediáticos poderosos para la generación de interés social en determinadas ideas políticas. Así, un medio sintético puede modificar lo que sabemos y alterar nuestras percepciones en torno a los agentes políticos.

En el quinto capítulo, “‘I don’t like them putting fraud in my conspiracy theories!’: a philosophical examination of conspiracy theories, ‘pataphysics, and Alex Jones’s Frog Conspiracy” por Javier Silva, Gabriel Donoso y Esteban Echaniz, se examina una supuesta teoría conspirativa

inofensiva para mostrar que, en efecto, constituye más bien un fraude con rasgos conspirativos. Este trabajo de aplicación muestra la necesidad de atender a las teorías conspirativas con precisión conceptual para dejar en evidencia si genuinamente pueden considerarse de bajo esa categoría por su modo de presentación y objetivos. En el caso específico analizado por los autores, además, se revela un aspecto ético preocupante del uso del discurso conspirativo: hay agentes que emplean esta clase de discursos para beneficiarse económicamente a costa de la ignorancia de otros.

Posteriormente, en el capítulo sexto “El auge de las revistas de baja calidad en el ámbito académico: causas, consecuencias y desafíos éticos” por Hernán Bendezú, se aborda un problema que permea una de las instancias más importantes para la solidificación y comunicación del conocimiento: la publicación académica orientada por fines sospechosos. De acuerdo con Bendezú, la publicación debe ser entendida como un proceso de desarrollo que requiere de ética, estrategia y formación continua. En este sentido, la elección de revistas de dudosa acreditación daría cuenta de una toma de riesgos innecesarios para la reputación de una investigación bien orientada, impidiendo su desarrollo en cuestiones centrales como el pensamiento crítico y el avance del conocimiento afianzado.

Finalmente, la sección cierra con “El entendimiento como un requisito para la confiabilidad institucional” por

Felipe Morales, en el que se examinan las condiciones que hacen confiables a las instituciones. En él, Morales defiende que las instituciones fiables deben tener un entendimiento adecuado de determinados problemas, manifestar dicho entendimiento y darse ellas mismas a entender haciéndose visibles reforzando, a su vez, las capacidades de las personas para entenderlas. De ser el caso contrario, no hay posibilidad para la confianza en las instituciones puesto que no se puede confiar en aquello que no se entiende. La sección cierra con esta contribución intencionadamente, pues permite reflexionar sobre una serie de fenómenos vistos a lo largo de ella (la posverdad, las *fake news*, la inteligencia artificial, los *deepfakes*, las teorías conspirativas, las publicaciones académicas) que beben, para bien o para mal, de la confianza social en las instituciones en las que ellas se apoyan o parasitan. En este sentido, la contribución de Morales nos da una pista clave para volver a examinar lo ya dicho: la importancia del entendimiento en relación con la confiabilidad institucional y todo lo que ello conlleva.

# Responsabilidades éticas en la era de la posverdad: desafíos epistémicos en las redes sociales digitales

*Ethical Responsibilities in the Post-Truth Era:  
Epistemic Challenges in Digital Social Networks*

**Rodolfo Aldea<sup>82</sup>**

## Resumen

En la era digital, la proliferación de la posverdad y las *fake news* plantea desafíos éticos y epistémicos que erosionan la confianza en el conocimiento colectivo. Este capítulo explora las responsabilidades morales de individuos, plataformas e instituciones en la difusión de información,

---

<sup>82</sup> Universidad Alberto Hurtado. <https://orcid.org/0009-0002-7506-8504>

integrando perspectivas éticas con análisis epistemológicos sobre la injusticia cognitiva. Se reflexiona sobre el rol de los algoritmos en la manipulación de la verdad y se proponen marcos para fomentar una comunicación digital responsable, considerando impactos políticos en democracias contemporáneas. El enfoque interdisciplinario busca contribuir a debates del siglo XXI sobre poder, verdad y sociedad, promoviendo herramientas para mitigar la desinformación y restaurar la integridad epistémica.

**Palabras clave:** Posverdad; ética digital; epistemología social; desinformación; redes sociales.

## **Abstract**

In the digital age, the proliferation of post-truth and fake news poses ethical and epistemic challenges that erode trust in collective knowledge. This chapter explores the moral responsibilities of individuals, platforms, and institutions in information dissemination, integrating ethical perspectives with epistemological analyses of cognitive injustice. It reflects on the role of algorithms in truth manipulation and proposes frameworks to foster responsible digital communication, considering political impacts on contemporary democracies. The interdisciplinary approach aims to contribute to 21st-century debates on power, truth, and society, promoting tools to mitigate disinformation and restore epistemic integrity.

**Keywords:** Post-truth; digital ethics; social epistemology; disinformation; social networks.

## **Introducción**

En el panorama contemporáneo, la posverdad emerge como un fenómeno que trasciende el mero engaño informativo para convertirse en un desafío estructural a la forma en que construimos y compartimos el conocimiento colectivo. Definida por algunos autores como una era en la que los hechos objetivos tienen menos influencia en la formación de la opinión pública que las apelaciones emocionales y las creencias personales (Kakutani, 2018), la posverdad se ha intensificado con la proliferación de las redes sociales digitales. Estas plataformas, diseñadas para maximizar el *engagement* a través de algoritmos que priorizan el contenido viral, facilitan la difusión rápida de información sesgada o falsa, erosionando la distinción entre verdad y ficción.

Este problema no es aislado, sino que se entrelaza con tendencias sociales más amplias, como la polarización ideológica y el declive en la confianza hacia instituciones tradicionales, incluyendo la ciencia y el periodismo. Por ejemplo, estudios recientes destacan cómo mega-tendencias como la desconfianza en expertos y la amplificación de narrativas alternativas contribuyen a un entorno donde la desinformación prospera (Lewandowsky, 2020). En este

contexto, fenómenos como las *fake news* y las teorías conspirativas no solo distorsionan el debate público, sino que también generan divisiones sociales profundas, afectando la cohesión democrática.

Desde una perspectiva crítica en estudios de comunicación, la posverdad representa una ruptura en la confianza social, donde el valor de la verdad se subordina a estrategias retóricas y emocionales (Harsin, 2018). En las redes sociales, esto se manifiesta en “burbujas informativas” (o “de filtro”<sup>83</sup>) que refuerzan prejuicios preexistentes, limitando la exposición a perspectivas contrarias y fomentando un ciclo de retroalimentación que perpetúa la desinformación. Así, la posverdad no es solo un problema epistemológico —relacionado con cómo sabemos lo que sabemos— sino también ético, al cuestionar las responsabilidades de quienes participan en la difusión digital del conocimiento.

Una distinción clave en este contexto es la propuesta por Nguyen (2020), quien diferencia entre “burbujas epistémicas”, donde voces relevantes se excluyen por omisión accidental o algorítmica (por ejemplo, por selección pasiva de contactos o *feeds* que priorizan contenido afín), y “cámaras de eco”, donde la exclusión es activa, en la medida en que se deslegitima sistemáticamente a fuentes externas, generando

---

<sup>83</sup> Ver Michiels et al. (2022) para un análisis crítico del concepto.



desconfianza profunda hacia todo lo ajeno al grupo. Esta diferenciación es crucial, ya que en las cámaras de eco la mera exposición a evidencia contraria no rompe el aislamiento — puede incluso reforzarlo al interpretarse como confirmación de la conspiración— mientras que las burbujas se disipan más fácilmente con acceso diverso a información. Esto agrega una variable más a la discusión.

El principal objetivo de este capítulo es examinar las responsabilidades éticas inherentes a la difusión del conocimiento en la era digital, con un énfasis particular en los desafíos epistémicos que surgen en las redes sociales. Específicamente, se busca explorar cómo la posverdad no solo altera la percepción de la verdad, sino que también impone obligaciones morales a individuos, plataformas y sociedades para mitigar sus efectos perjudiciales. A través de este análisis, se pretende contribuir a los debates filosófico-políticos contemporáneos, proponiendo marcos que fomenten una comunicación digital más responsable y equitativa.

Para lograr esto, el capítulo adopta un enfoque interdisciplinario que integra la ética de la información con la epistemología social. Desde la ética, se parte de la premisa de que el conocimiento no es neutral, sino que conlleva responsabilidades en su manejo y distribución, especialmente en entornos donde la información puede ser manipulada para fines personales o políticos (Floridi, 2013). Esto se

complementa con perspectivas epistemológicas que cuestionan cómo sabemos y validamos el conocimiento en contextos colectivos, reconociendo que las redes sociales alteran las dinámicas tradicionales de creencia y justificación (O'Connor et al., 2024).

Esta integración permite un análisis holístico en la medida en que no solo se describen los problemas, sino que se reflexiona sobre sus implicaciones para la justicia cognitiva y el poder democrático. De esta manera, se busca no solo diagnosticar, sino también ofrecer herramientas conceptuales para navegar los dilemas del siglo XXI.

La relevancia de estos desafíos trasciende el ámbito individual para insertarse en debates filosófico-políticos más amplios del siglo XXI, donde la posverdad amenaza las bases mismas de la democracia y el orden social. En un mundo interconectado por redes digitales, la manipulación de la información no solo distorsiona el conocimiento, sino que también redefine las estructuras de poder, permitiendo que actores estatales, corporativos o ideológicos influyan en el imaginario colectivo de manera inédita (Habermas, 2018). Esto plantea interrogantes éticos sobre la soberanía informativa y la equidad en el acceso a la verdad, especialmente en contextos donde las desigualdades digitales exacerban divisiones sociales.

Desde una perspectiva política contemporánea, la posverdad contribuye a una crisis democrática al fomentar la polarización y el populismo, donde las narrativas alternativas socavan el consenso racional necesario para la deliberación pública. Autores destacan cómo este fenómeno representa un riesgo para las instituciones liberales, al erosionar la confianza en procesos electorales y políticas públicas basadas en evidencia (Farkas & Schou, 2020). En este sentido, reflexionar sobre responsabilidades éticas y epistémicas no es un ejercicio abstracto, sino una herramienta para reimaginar formas de gobernanza que integren la sostenibilidad digital y la justicia cognitiva.

Así, este capítulo se posiciona en el cruce de la filosofía política y la ética aplicada, contribuyendo a un diálogo urgente sobre cómo navegar un mundo post-verdad sin sacrificar los principios de verdad, responsabilidad y equidad que definen la convivencia humana en el siglo XXI.

## **Fundamentos éticos en la difusión del conocimiento digital**

En la era de las redes sociales, la responsabilidad ética en la difusión del conocimiento se extiende tanto a los individuos como a las colectividades, transformando prácticas cotidianas en actos con implicaciones morales profundas. Los usuarios no son meros consumidores pasivos, sino agentes activos que, al compartir, comentar o amplificar

información, contribuyen a la construcción de narrativas colectivas. Esta responsabilidad individual implica un deber de diligencia, esto es, verificar fuentes antes de diseminar contenido, evitando así la propagación inadvertida de desinformación que podría dañar la confianza pública o incitar divisiones sociales (Wardle & Derakhshan, 2017).

A nivel colectivo, las plataformas digitales actúan como intermediarios que moldean el flujo informativo, lo que plantea obligaciones éticas para diseñadores y moderadores. Estas entidades deben priorizar la transparencia y la equidad, reconociendo que sus algoritmos no solo reflejan, sino que también amplifican sesgos humanos. Por ejemplo, en contextos juveniles, las redes sociales complican la formación de identidades al exponer a usuarios a flujos informativos manipulados, demandando una ética que fomente entornos seguros y educativos (Boyd, 2014). Desde una perspectiva ética aplicada, esto implica adoptar principios como la no-maleficencia y la justicia, asegurando que la difusión del conocimiento no perpetúe desigualdades.

Además, la responsabilidad colectiva se entrelaza con deberes cívicos en democracias digitales, donde la inacción ante la desinformación puede erosionar el tejido social. Autores enfatizan que las redes no son neutrales, sino espacios donde las interacciones éticas determinan la calidad del discurso público, instando a usuarios y plataformas a cultivar virtudes como la empatía y el escepticismo crítico

(Vallor, 2022). De esta forma, la responsabilidad no es solo reactiva, sino proactiva, orientada a fomentar un ecosistema digital donde el conocimiento sirva al bien común.

Pero el desafío ético no se agota en la conducta humana. Uno de los actores más influyentes —aunque invisibles— son los algoritmos que estructuran nuestras interacciones digitales. Los algoritmos que rigen las redes sociales representan un pilar central en la difusión del conocimiento digital, pero su diseño plantea dilemas éticos profundos relacionados con la manipulación implícita y los sesgos inherentes. Estos sistemas, optimizados para maximizar el tiempo de usuario y la interacción, a menudo priorizan contenido sensacionalista o “polarizante”, lo que distorsiona la presentación de información y fomenta entornos donde la verdad se subordina a la viralidad (Mittelstadt et al., 2016). Esta manipulación no es siempre intencional, pero sus consecuencias éticas son significativas, en cuanto algoritmos opacos pueden amplificar prejuicios humanos codificados en datos de entrenamiento, perpetuando desigualdades raciales, de género o ideológicas en la recomendación de contenido.

Desde una perspectiva ética, esto viola principios como la autonomía y la no-discriminación, al limitar la exposición diversa y crear “burbujas de filtro” que aíslan a los usuarios en ecosistemas informativos homogéneos (Pariser, 2011). En el contexto de la posverdad, estos sesgos facilitan la

propagación de *fake news*, ya que algoritmos premian el *engagement* emocional sobre la veracidad factual. Autores critican este modelo como una forma de “capitalismo de vigilancia”, donde el conocimiento se mercantiliza para fines comerciales, erosionando la integridad ética de la información pública (Zuboff, 2023).

Para abordar estos desafíos, es esencial demandar transparencia algorítmica y auditorías éticas, reconociendo que los algoritmos no son neutrales, sino artefactos morales que deben alinearse con valores sociales como la justicia y la responsabilidad. De lo contrario, su manipulación implícita no solo distorsiona el conocimiento, sino que también socava la confianza colectiva en la era digital.

Esta dimensión ética se entrelaza estrechamente con implicaciones políticas más amplias, especialmente cuando las tecnologías digitales actúan como instrumentos de poder y control. Las intersecciones entre ética y política en la difusión del conocimiento digital revelan cómo las redes sociales se convierten en arenas de poder y control, donde la información no es solo un bien público, sino una herramienta para la dominación. En entornos digitales, los algoritmos y plataformas ejercen un control sutil sobre qué se ve y se cree, reconfigurando dinámicas políticas tradicionales hacia formas de tecnopolítica que priorizan la manipulación sobre el diálogo racional (Habermas, 2018). Esto plantea dilemas éticos sobre la concentración de poder en manos de

corporaciones privadas, que pueden influir en elecciones, movimientos sociales y opiniones públicas sin rendir cuentas adecuadas.

Por ejemplo, en democracias divididas, las redes amplifican propaganda y desinformación que radicalizan audiencias, fomentando un ecosistema donde el poder se ejerce a través de narrativas virales en lugar de instituciones formales (Benkler et al., 2018). Desde una perspectiva ética, esto viola principios de equidad política, al crear asimetrías donde grupos marginados enfrentan mayor exposición a contenido manipulador, exacerbando desigualdades. Autores advierten que esta "república dividida" en la era de las redes sociales erosiona la deliberación democrática, sustituyéndola por "cámaras de eco" que priorizan la lealtad tribal sobre la verdad compartida (Sunstein, 2017).

En última instancia, abordar estas intersecciones requiere una ética política que integre regulaciones para limitar el control algorítmico, promoviendo transparencia y responsabilidad como pilares para un poder digital más justo. Sin esto, los entornos digitales arriesgan perpetuar ciclos de control que socavan la autonomía colectiva y la integridad ética de la sociedad.

## Desafíos epistémicos en la posverdad

En la era de la posverdad, las formas de desinformación como las *fake news*, los rumores y las teorías conspirativas representan desafíos epistémicos fundamentales, al distorsionar la base misma del conocimiento compartido en redes sociales. Las *fake news*, definidas como información fabricada intencionalmente para engañar, se propagan a velocidades alarmantes gracias a algoritmos que favorecen el contenido emocional y controvertido, superando a menudo la difusión de hechos verificados (Vosoughi et al., 2018). Esto no solo confunde a los usuarios, sino que también erosiona la credibilidad de fuentes confiables, creando un entorno donde la verdad se relativiza.

Los rumores, por su parte, operan como formas patológicas de testimonio, amplificadas en ecosistemas digitales donde la verificación es escasa y la viralidad prima sobre la precisión. Estos se nutren de la incertidumbre social, evolucionando rápidamente en narrativas que refuerzan prejuicios preexistentes y fomentan desconfianza colectiva (Coady, 2012). Similarmente, las teorías conspirativas prosperan en burbujas epistémicas, donde algoritmos curan *feeds* que aíslan a los individuos en comunidades cerradas, limitando la exposición a contraevidencias y perpetuando creencias infundadas (Badman, 2022).



Desde una perspectiva epistemológica, estas formas de desinformación generan un "dividido epistémico" digital, donde el acceso desigual a información veraz agrava desigualdades cognitivas y socava la formación racional de creencias (McIntyre, 2018). En redes sociales, el problema se agrava por la ausencia de *gatekeepers* tradicionales, permitiendo que la desinformación se convierta en una herramienta de manipulación masiva, con impactos que van desde la salud pública hasta la estabilidad política.

Más allá del contenido falso, la posverdad afecta también quién tiene el derecho a ser creído. Aquí entra en juego la noción de injusticia epistémica. Dicha noción, entendida como el daño infligido a alguien en su capacidad como conocedor, se manifiesta de manera aguda en la era de la posverdad, particularmente en relación con la desconfianza hacia la ciencia y el *expertise*. Este fenómeno no solo involucra la deslegitimación de voces autorizadas, sino que también perpetúa desigualdades cognitivas donde ciertos grupos —como minorías étnicas, mujeres o comunidades marginadas— ven sus testimonios sistemáticamente desvalorizados o ignorados en plataformas digitales (Fricker, 2007). En redes sociales, algoritmos que priorizan contenido controvertido amplifican voces anti-científicas, creando un entorno donde la evidencia empírica se equipara a opiniones infundadas, erosionando la confianza colectiva en instituciones científicas.

Esta desconfianza se agrava por la posverdad, donde narrativas emocionales suplantán hechos verificables, generando un “déficit hermenéutico” —como lo plantea Fricker (2007)— que impide a grupos históricamente excluidos articular sus experiencias de injusticia cognitiva (véase también Sugeng et al., 2024). Este déficit hermenéutico, entendido como la carencia de marcos interpretativos adecuados para comprender y comunicar ciertas experiencias, se manifiesta de manera aguda en contextos atravesados por desigualdades de género, clase y etnicidad. Según Kristie Dotson (2012), la injusticia epistémica puede adquirir formas estructurales más profundas cuando las comunidades marginadas no solo carecen de acceso a los recursos interpretativos dominantes, sino que además sus aportes son sistemáticamente desestimados. En estos casos, el silenciamiento epistémico no es accidental, sino producto de contextos históricos de exclusión que limitan la posibilidad de que dichas comunidades definan los términos de su propia realidad. Así, el déficit hermenéutico en la era digital no solo restringe el reconocimiento de estas voces, sino que perpetúa desigualdades cognitivas al reforzar narrativas que privilegian perspectivas dominantes.

En otros casos, como durante crisis globales en pandemias, la desinformación sobre vacunas no solo fomenta escepticismo, sino que también refleja una mutación epistemológica en el periodismo y la comunicación digital,

donde la verdad se fragmenta en versiones alternativas que socavan el consenso científico (Habermas & Vallespín, 2021). Autores señalan que internet, al democratizar el acceso al conocimiento, paradójicamente lo debilita al exponer a usuarios a flujos informativos que priorizan la cantidad sobre la calidad, resultando en una comprensión superficial y fragmentada (Lynch, 2016). Un ejemplo de ello se observa en la circulación masiva de hilos de redes sociales que condensan investigaciones científicas complejas en listados breves o infografías simplificadas, los que, si bien facilitan su consumo rápido, eliminan matices esenciales y, en ocasiones, distorsionan las conclusiones originales, fomentando la ilusión de comprensión sin un verdadero análisis crítico.

Desde una lente epistemológica social, esta injusticia no es accidental, sino estructural, i.e., las redes perpetúan sesgos que desacreditan expertos en favor de *influencers* o narrativas populistas, exacerbando divisiones entre legos y especialistas. Esto no solo distorsiona el conocimiento científico, sino que también tiene implicaciones éticas, al privar a sociedades de herramientas racionales para enfrentar desafíos globales como el cambio climático o la salud pública.

Las dinámicas de desacuerdo en la era de la posverdad representan un desafío epistémico clave, particularmente en las redes sociales donde las interacciones entre pares epistémicos —individuos con niveles similares de conocimiento— y entre legos y expertos se fragmentan y

polarizan. Entre pares, el desacuerdo surge cuando creencias opuestas se atrincheran en burbujas informativas, donde algoritmos refuerzan posiciones preexistentes en lugar de fomentar el intercambio racional, resultando en un estancamiento cognitivo que prioriza la lealtad grupal sobre la evidencia compartida (O'Connor et al., 2024). Esto no solo perpetúa la desinformación, sino que también erosiona la capacidad colectiva para resolver disputas a través de deliberación.

En el caso de desacuerdos entre legos y expertos, la posverdad agrava la brecha al deslegitimar el *expertise*, equiparando opiniones no informadas con conocimiento especializado. Plataformas digitales amplifican voces amateurs que cuestionan consensos científicos, como en debates sobre cambio climático o vacunas, generando un "dividido epistémico digital" donde la confianza en expertos se erosiona por narrativas populistas o conspirativas (Abiri, et al., 2022). Desde una perspectiva epistemológica social, esto refleja una falla en los mecanismos de justificación: legos, expuestos a información sesgada, rechazan evidencia experta no por falta de acceso, sino por filtros algorítmicos que privilegian el escepticismo emocional.

Estas dinámicas no son neutras; implican injusticias éticas al marginar voces expertas y empoderar desinformación, afectando decisiones políticas y sociales. Para mitigarlas, es esencial promover normas epistémicas

que incentiven el diálogo cruzado, reconociendo que en entornos digitales, el desacuerdo no resuelto puede escalar a divisiones irreconciliables que socavan la cohesión democrática.

## **Propuestas para una comunicación responsable**

Para contrarrestar los desafíos éticos y epistémicos de la posverdad, es imperativo desarrollar marcos éticos robustos que guíen tanto a las plataformas digitales como a los usuarios en la difusión responsable del conocimiento. Estos marcos deben basarse en principios de transparencia, *accountability* y justicia, reconociendo que la información digital no es un *commodity* neutral, sino un recurso con impactos morales (Floridi, 2013). Para las plataformas, un enfoque clave implica la implementación de auditorías algorítmicas independientes que evalúen sesgos y promuevan diversidad en los *feeds*, asegurando que el diseño priorice la veracidad sobre el *engagement* puramente comercial.

Un ejemplo de marco propuesto es un "*charter* global para la verdad" adaptado a la era de internet, que establece normas para combatir la desinformación mediante herramientas como verificación integrada y etiquetado de contenido dudoso, fomentando una epistemología social inclusiva (Habgood-Coote, 2021). Para los usuarios, estos marcos enfatizan la educación en "literacidad digital ética", instando a prácticas como la reflexión crítica antes de

compartir y el reconocimiento de responsabilidades individuales en la cadena informativa. En contextos globales, se propone mejorar plataformas como X (ex Twitter) mediante diseños que mitiguen injusticias epistémicas, promoviendo *engagement* inclusivo que amplifique voces marginadas y reduzca polarización (Chavanayarn, 2024).

Estos marcos no son prescriptivos rígidos, sino flexibles, integrando *feedback* comunitario para adaptarse a evoluciones tecnológicas. Al adoptarlos, plataformas y usuarios pueden transformar las redes en espacios de comunicación responsable, restaurando la confianza y alineándose con valores éticos que prioricen el bien común sobre intereses particulares.

La educación y la regulación emergen como pilares fundamentales para restaurar la confianza en la comunicación digital, abordando los vacíos éticos y epistémicos que la posverdad ha exacerbado. En el ámbito educativo, es crucial implementar programas de literacidad digital que no solo enseñen habilidades técnicas, como la verificación de fuentes, sino que también fomenten virtudes éticas como la empatía y el pensamiento crítico en entornos en red. Por ejemplo, iniciativas dirigidas a jóvenes, que son nativos digitales, pueden ayudar a navegar complejidades sociales en plataformas, promoviendo un uso responsable que mitigue la propagación de desinformación y fortalezca la resiliencia cognitiva (Boyd, 2014).

La regulación, por su parte, debe equilibrar la libertad de expresión con protecciones contra abusos, mediante marcos legales que exijan transparencia en algoritmos y responsabilidad de plataformas por contenido dañino. Esto incluye políticas que aborden brechas epistémicas entre periodistas y audiencias, donde la desinformación crea expectativas divergentes sobre la veracidad, instando a regulaciones que incentiven la colaboración entre medios y usuarios para reconstruir confianza (Godler & Reich, 2024). En un enfoque interdisciplinario, la educación y regulación se complementan: la primera empodera a individuos, mientras la segunda impone estándares colectivos, como multas por negligencia en moderación o incentivos para educación pública en escuelas y comunidades.

Al integrar estos elementos, se puede fomentar una cultura de confianza restaurada, donde la comunicación responsable no sea opcional, sino un estándar ético que contrarreste la erosión de la verdad en la era digital, contribuyendo a sociedades más informadas y cohesionadas.

En el marco de una comunicación responsable, las reflexiones sobre la democracia digital y la sostenibilidad epistémica destacan la necesidad de reimaginar las redes sociales como espacios que fortalezcan, en lugar de debilitar, los pilares democráticos. La democracia digital, caracterizada por la participación masiva en plataformas en línea, enfrenta riesgos cuando la posverdad fragmenta el espacio público en

"repúblicas divididas", donde el diálogo racional cede ante polarizaciones algorítmicas que priorizan el conflicto sobre el consenso (Sunstein, 2017). Para sostenerla, es esencial promover diseños que incentiven la deliberación inclusiva, integrando herramientas éticas que mitiguen burbujas informativas y fomenten exposiciones cruzadas a perspectivas diversas.

La sostenibilidad epistémica, por su parte, se refiere a la capacidad a largo plazo de mantener ecosistemas de conocimiento resilientes frente a amenazas como la desinformación persistente. Esto implica una crítica normativa de las redes, donde se cuestiona su rol en la erosión de normas epistémicas compartidas, proponiendo fines que prioricen la verdad colectiva sobre el individualismo cognitivo (Habermas, 2020). En contextos de populismo epistémico, donde la desconfianza en hechos se convierte en herramienta política, la sostenibilidad requiere enfoques que restauren la presentación confiable de información, evitando que narrativas falsas dominen el discurso público (Nawrocki, 2023).

Estas reflexiones convergen en una visión holística: una democracia digital sostenible debe anclarse en principios éticos que regulen el poder algorítmico, eduquen a ciudadanos en virtudes epistémicas y fomenten regulaciones globales para la equidad informativa. De esta manera, las redes pueden evolucionar hacia plataformas que no solo



conecten, sino que también preserven la integridad del conocimiento democrático, asegurando un futuro donde la posverdad no dicte las reglas de la convivencia.

No obstante, cualquier marco normativo o ético propuesto para contrarrestar la posverdad debe ser sensible a los contextos culturales, políticos y socioeconómicos donde se implementa. Las dinámicas de desinformación y las formas de injusticia epistémica no se manifiestan de manera uniforme en todos los entornos; por el contrario, adoptan características particulares en función de factores históricos, institucionales y tecnológicos. En este sentido, resulta crucial considerar cómo estos fenómenos se desarrollan en regiones como América Latina, donde las brechas digitales, la polarización política y la instrumentalización mediática configuran un escenario especialmente vulnerable a los efectos de la posverdad.

### *Consideraciones para contextos latinoamericanos*

En América Latina, el fenómeno de la posverdad adquiere una complejidad singular al insertarse en democracias frágiles, profundamente marcadas por desigualdades estructurales, crisis institucionales y la desconfianza ciudadana hacia los medios y los actores políticos. Como señalan Flórez Hernández y Marlés Herrera (2022), la región vive no solo una saturación informativa, sino también una "constante información, desinformación y *fake*

*news* que permea las mentes y erosiona la libertad de decisión y acción en los procesos democráticos" (p. 85).

La posverdad en este contexto no es solo un fenómeno de orden epistémico, sino una herramienta de poder profundamente política. Las noticias falsas se utilizan estratégicamente en procesos electorales y movilizaciones sociales, generando entornos de manipulación que desactivan el pensamiento crítico. Esto se agrava con la dificultad generalizada de identificar la veracidad de las noticias: el 70% de los latinoamericanos no sabe reconocer una *fake news* en Internet, según el estudio *Digital Iceberg* citado en el mismo artículo (p. 88).

Además, la posverdad se articula como un arma discursiva: los líderes políticos desacreditan la información que les resulta incómoda, etiquetándola de *fake news*, incluso cuando se trata de hechos comprobables. Esto ha sido evidente en países como El Salvador, Brasil, Argentina o Colombia, donde las redes sociales han sido instrumentalizadas para reforzar narrativas populistas o polarizantes. En este marco, la posverdad no solo distorsiona la deliberación pública, sino que también consolida un modelo de "democracia espectador" (Han, 2015), en la que los ciudadanos actúan como consumidores pasivos de información "emocionalizada".

Este fenómeno se amplifica en un entorno mediático dominado por la deslegitimación de los medios tradicionales, la propagación de teorías conspirativas y la circulación viral de contenidos falsos. Según Flórez Hernández y Marlés Herrera (2022), este entorno configura una "dictadura de la emoción" (p. 96), donde las decisiones políticas son impulsadas por reacciones afectivas antes que por deliberaciones racionales. Esto compromete no solo la calidad del debate público, sino el ejercicio mismo de la libertad individual, al estar la voluntad del sujeto influida por estructuras comunicativas opacas y manipuladoras.

Un caso particular ilustrativo en América Latina se observa en las elecciones presidenciales argentinas de 2023, donde, como analiza Łabuz y Nehring (2024), ambos candidatos recurrieron extensivamente a contenido generado por IA y a *deepfakes*, lo que llevó a describir el proceso como "la primera elección con IA" (Nicas, 2023, citado en Łabuz & Nehring, 2024). La campaña de Sergio Massa empleó esta tecnología a gran escala para generar pósters electorales estilizados (algunos como propaganda soviética clásica o referencias pop), pero también para desacreditar a Javier Milei. Por ejemplo, difundieron un video *deepfake* en el que Milei explicaba supuestamente cómo funcionaría el negocio de la venta de órganos humanos (marcado como IA, pero diseñado para dañar su reputación), junto con otros materiales que lo presentaban como zombie o loco.

En respuesta, la campaña de Milei compartió imágenes generadas por IA que retrataban a Massa como un líder comunista en atuendo militar, alcanzando hasta 30 millones de vistas en redes sociales. Estos contenidos, mayoritariamente paródicos o satíricos se viralizaron en plataformas como X e Instagram, incentivando a simpatizantes de ambos lados a experimentar y producir material similar, lo que generó una dinámica de “*arms race*” en la manipulación digital. Esta instrumentalización exacerbó la “dictadura de la emoción” (Flórez Hernández & Marlés Herrera, 2022) y erosionó la integridad epistémica al normalizar la manipulación algorítmica, donde grabaciones reales fueron etiquetadas como falsas por partidarios. Así, este caso ilustra cómo la posverdad se convierte en herramienta de poder en democracias frágiles, perpetuando injusticias cognitivas al deslegitimar hechos verificables y amplificar polarizaciones que socavan la deliberación racional.

En suma, abordar el problema de la posverdad en América Latina exige una mirada crítica que reconozca sus raíces estructurales y sus implicancias éticas y políticas. La lucha contra las *fake news* no puede reducirse a una cuestión técnica de verificación, sino que requiere fortalecer las capacidades ciudadanas, recuperar la confianza en las instituciones y establecer mecanismos éticos de responsabilidad informativa tanto para los medios como para los actores políticos y sociales.

## Conclusiones

En síntesis, los debates ético-epistémicos explorados en este capítulo revelan cómo la posverdad en las redes sociales digitales no es un mero accidente informativo, sino un desafío estructural que interconecta responsabilidades morales con la integridad del conocimiento colectivo. Desde el plano ético, hemos visto que la difusión del conocimiento impone deberes individuales y colectivos, donde algoritmos y usuarios deben priorizar la veracidad para evitar manipulaciones que erosionen la autonomía y la justicia (Fricker, 2007). Epistémicamente, fenómenos como la desinformación, la injusticia cognitiva y los desacuerdos polarizados destacan la fragilidad de la confianza en entornos digitales, donde mega-tendencias sociales amplifican la polarización y el escepticismo hacia la ciencia (Lewandowsky, 2020).

Estos debates convergen en la necesidad de un enfoque interdisciplinario que reconozca la posverdad como una erosión de normas compartidas, afectando no solo la percepción individual, sino la cohesión social. Al integrar ética y epistemología social, se evidencia que, sin intervenciones responsables, las redes perpetúan ciclos de desconfianza que socavan la base racional de la sociedad. Esta síntesis subraya que la resolución no radica en soluciones técnicas aisladas, sino en marcos holísticos que restauren el valor ético de la verdad en la era digital.

Las implicaciones de estos debates para el pensamiento político contemporáneo son profundas, ya que la posverdad en entornos digitales redefine las estructuras de poder y la gobernanza democrática. En un mundo donde las redes sociales actúan como foros públicos dominantes, la manipulación de la información puede exacerbar la propaganda y la radicalización, socavando el consenso racional necesario para políticas inclusivas y basadas en evidencia (Benkler et al., 2018). Esto invita a repensar el legado de teorías clásicas, integrando conceptos de tecnopolítica que aborden cómo el control algorítmico influye en la formación de opiniones y en procesos electorales, fomentando un paradigma donde la sostenibilidad digital sea central.

Desde una perspectiva crítica, la posverdad representa una crisis democrática al limitar el discurso público y promover divisiones, cuestionando la viabilidad de democracias liberales en un contexto de desinformación generalizada (Farkas & Schou, 2020). En estudios de comunicación, se enfatiza que esta ruptura en la confianza social demanda un replanteamiento político que priorice la regulación ética y la educación cívica para restaurar la integridad del debate público (Harsin, 2018). Así, el pensamiento político del siglo XXI debe incorporar estas reflexiones, promoviendo modelos híbridos que equilibren libertad digital con responsabilidades colectivas, para evitar

que la posverdad dicte las dinámicas de poder y equidad social.

Las sugerencias para investigaciones futuras deben enfocarse en expandir los debates ético-epistémicos más allá de diagnósticos generales, hacia análisis empíricos y normativos que aborden contextos específicos en la era digital. Por ejemplo, estudios comparativos sobre cómo la posverdad afecta regiones como América Latina podrían integrar perspectivas culturales locales, explorando cómo redes sociales influyen en dinámicas políticas regionales y proponiendo marcos adaptados a desigualdades digitales persistentes (Lynch, 2016). Esto podría incluir investigaciones interdisciplinarias que combinen epistemología social con estudios de comunicación, evaluando el impacto de intervenciones educativas en la mitigación de injusticias epistémicas.

Además, futuras indagaciones podrían examinar el rol emergente de tecnologías como la IA en la amplificación de desinformación, desarrollando modelos éticos para regular su uso en plataformas y midiendo su efectividad en restaurar confianza (McIntyre, 2018). Otro frente prometedor es el análisis longitudinal de desacuerdos epistémicos en eventos reales, como elecciones o crisis sanitarias, para identificar patrones que informen políticas públicas sostenibles. En última instancia, estas investigaciones deberían priorizar enfoques colaborativos globales, incorporando voces de

disciplinas afines para enriquecer el pensamiento filosófico-político y enfrentar los desafíos del siglo XXI con herramientas innovadoras y equitativas.

## Referencias

- Abiri, G., & Buchheim, J. (2022). Beyond true and false: Fake news and the digital epistemic divide. *Michigan Technology Law Review*, 29, 59.
- Badman, R. (2022). *Bubbles and chambers: Post-truth and belief formation*. PhilArchive.
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- <https://doi.org/10.1093/oso/9780190923624.001.0001>
- boyd, d. (2014). *It's complicated: The social lives of networked teens*. Yale University Press.
- <https://doi.org/10.12987/9780300166439>
- Chavanayarn, S. (2024). Epistemic injustice and ideal social media: Enhancing X for inclusive global engagement. *Topoi*. <https://doi.org/10.1007/s11245-024-10070-7>
- Coady, D. (2012). *What to believe now: Applying epistemology to contemporary issues*. John Wiley & Sons.
- Dotson, K. (2012). A cautionary tale: On limiting epistemic oppression. *Frontiers: A Journal of Women Studies*,



33(1), 24–47.

<https://doi.org/10.5250/fronjwomestud.33.1.0024>

Farkas, J., & Schou, J. (2020). Post-truth discourses and their limits: A democratic crisis? *International Journal of Communication*.

Flórez Hernández, C. A., & Marlés Herrera, J. A. (2022). Fake news and democracy in Latin America: A new paradigm in the political–media relationship. *Cuestiones Políticas*, 40(72), 82–98.

<https://doi.org/10.46398/cuestpol.4072.05>

Floridi, L. (2013). *The ethics of information*. Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780199641321.001.0001>

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>

Godler, Y., & Reich, Z. (2024). The threat of misinformation on journalism's epistemology: Exploring the gap between journalist's and audience's expectations. *Digital Journalism*.

<https://doi.org/10.1080/21670811.2024.2320249>

Habermas, J. (2018). *The lure of technocracy*. Polity Press.

Habermas, J. (2020). *Towards a critical social epistemology of social media*. PhilArchive.

Habermas, J., & Vallespín, F. (2021). Post-truth as a mutation of epistemology in journalism. *Media and Communication*.

<https://doi.org/10.17645/mac.v9i1.3529>

Habgood-Coote, J. (2021). *A global charter for truth*.

Han, B.-C. (2015). *La sociedad de la transparencia*. Herder.

Harsin, J. (2018). Post-truth and critical communication studies. En *Oxford Research Encyclopedia of Communication*. Oxford University Press. Recuperado 7 de agosto de 2025, de <https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-757>

Kakutani, M. (2018). *The death of truth: Notes on falsehood in the age of Trump*. Tim Duggan Books.

Łabuz, M., & Nehring, C. (2024). On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *European Political Science*, 23(4), 454–473. <https://doi.org/10.1057/s41304-024-00482-9>

Lewandowsky, S. (2020). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*.

Lynch, M. P. (2016). *The internet of us: Knowing more and understanding less in the age of big data*. Liveright.

McIntyre, L. (2018). *Post-truth in social media*. ResearchGate.

- Michiels, L., Leysen, J., Smets, A., & Goethals, B. (2022, julio). What are filter bubbles really? A review of the conceptual and empirical work. En *Adjunct proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 274–279).
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
- Nawrocki, S. (2023). Post-truth politics and epistemic populism: About (dis-)trusted presentation and communication of facts, not false information. Springer. [https://doi.org/10.1007/978-3-031-64178-7\\_2](https://doi.org/10.1007/978-3-031-64178-7_2)
- Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161. <https://doi.org/10.1017/epi.2018.32>
- O'Connor, C., Goldberg, S., & Goldman, A. (2024). Social epistemology. En E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Edición de Verano 2024). <https://plato.stanford.edu/archives/sum2024/entries/epistemology-social/>
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.

- Sugeng, S., Fitria, A., & Bastomi, S. (2024). Epistemic injustice and digital disinformation: Addressing knowledge inequities in the digital age. *Diskursus: Jurnal Filsafat dan Teologi STF Driyarkara*, 20(1), 135–169.
- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.  
<https://doi.org/10.1515/9781400884711>
- Vallor, S. (2022). Social networking and ethics. En E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Edición de Otoño 2022).  
<https://plato.stanford.edu/archives/fall2022/entries/ethics-social-networking/>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.  
<https://doi.org/10.1126/science.aap9559>
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe.
- Zuboff, S. (2023). The age of surveillance capitalism. En *Social theory re-wired* (pp. 203–213). Routledge.

# **El conocimiento como bien común: una reflexión ética sobre su difusión en la era de la inteligencia artificial**

*Knowledge as a Common Good: An Ethical Reflection on Its Dissemination in the Age of Artificial Intelligence*

**David Eduardo León<sup>84</sup>**

## **Resumen**

Este capítulo analiza los dilemas éticos asociados a la difusión del conocimiento en un contexto global marcado por el desarrollo de tecnologías de inteligencia artificial. A partir de debates contemporáneos en filosofía política,

---

<sup>84</sup> Servicio Nacional de Aprendizaje SENA. <https://orcid.org/0000-0002-7339-0068>

epistemología social y estudios críticos de tecnología, se examina la relación entre poder epistémico, justicia cognitiva y mediación algorítmica. El texto aborda conceptos como responsabilidad epistémica, injusticia algorítmica y equidad cognitiva, evaluando cómo sistemas de inteligencia artificial influyen en el acceso, producción y validación del conocimiento. Asimismo, se analiza la tensión entre el conocimiento como bien común y su creciente mercantilización en el ecosistema académico digital. Finalmente, se propone una ética digital del conocimiento basada en principios de transparencia algorítmica, pluralidad epistémica y acceso abierto, entendidos como condiciones necesarias para una circulación más justa del saber en las sociedades contemporáneas.

**Palabras clave:** Conocimiento abierto; inteligencia artificial; responsabilidad epistémica; justicia cognitiva; ética digital.

## **Abstract**

This chapter examines current ethical dilemmas in the dissemination of knowledge within a global context shaped by artificial intelligence technologies. Drawing on contemporary debates in political philosophy, social epistemology, and critical studies of technology, it explores the relationship between epistemic power, cognitive justice, and algorithmic mediation. The chapter discusses epistemic responsibility,

algorithmic injustice, and cognitive equity, assessing how AI systems influence access to, production of, and validation of knowledge. It also analyzes the tension between knowledge as a common good and its increasing commodification in today's academic and digital ecosystems. The chapter concludes by proposing a digital ethics of knowledge grounded in algorithmic transparency, epistemic plurality, and open access as necessary conditions for a more just circulation of knowledge.

**Keywords:** Open knowledge; artificial intelligence; epistemic responsibility; cognitive justice; digital ethics.

## **Introducción**

El conocimiento a lo largo de la historia ha sido un recurso para el debate. Desde las antiguas bibliotecas de Mesopotamia, pasando por los scriptoriums medievales, las academias ilustradas hasta los actuales centros de investigación científica y los repositorios digitales, el conocimiento ha estado íntimamente ligado a estructuras de poder, a la legitimidad del mismo y al acceso al conocimiento. La modernidad, por su parte, prometió un conocimiento más libre, más racional, más democrático, pero esa promesa, llegados a este momento, está condicionada por intereses políticos, económicos y tecnológicos que, en parte, mediatizan el conocimiento en quién lo tiene, acerca de quién lo produce, y para qué se distribuye.

Por otro lado, el desarrollo de nuevas tecnologías de inteligencia artificial (IA) y su uso masivo en motores de búsqueda, redes sociales, bibliotecas digitales o plataformas educativas han supuesto una modificación en las formas de circulación y dispersión del saber, al haberse desplazado lo que antes pasaba por los intermediarios humanos (editores, bibliotecarios, docentes, etc.) a depender, en buena parte, de los algoritmos que se ocupan de seleccionar, priorizar y recomendar contenido a partir de criterios a menudo invisibles para los usuarios. Esta lógica no solo incide en lo que aprendemos, sino también en lo que creemos que es verdadero, importante o fiable.

El hecho de abordar la ética en la difusión del conocimiento en este contexto debe significar no solo revisar los contenidos que circulan, sino que deberá también implicar el análisis de los dispositivos técnico-políticos que los sostienen. La inteligencia artificial no se presenta como un simple artefacto neutro, sino que se alimenta de datos, es guiada por objetivos y es configurada políticamente por intereses humanos. Como ha afirmado Turilli & Floridi (2009), hoy, la información es una de las dimensiones ontológicas de nuestra existencia, y por ello su tratamiento ético debe incluir los flujos de datos y también las estructuras que la hacen posible.

Desde la filosofía política, el conocimiento ha sido considerado por autores como Foucault (1972) como una



forma de poder que se distribuye de manera desigual, no solamente entre clases sociales, sino también entre culturas, géneros, generaciones y geografías. Las herramientas tecnológicas que median el acceso a la información, lejos de neutralizar desigualdades, pueden reproducirlas y/o incrementarlas. Las injusticias epistémicas son, tal como ha señalado Zobel (2019), no solo actos individuales de exclusión, sino también estructuras sistemáticas que silencian voces, restringen miradas y marginan formas alternativas de saber.

Para poder enfrentar este reto que se nos presenta, este capítulo prevé una reflexión filosófico-ética que le dé consistencia a la noción de responsabilidad epistémica en un escenario social marcado por la inteligencia artificial. La idea que propone es que los sistemas de inteligencia artificial, al coadyuvar directamente en qué conocimiento se difunde y cómo se accede a él, deben ser evaluados en base a parámetros no solamente técnicos, sino también a partir de sus capacidades en promover una equidad cognitiva, una pluralidad epistémica y el derecho al conocimiento. El presente análisis toma lugar en el contexto de una perspectiva que considera el conocimiento como un bien común tal como lo defendieron Hess y Ostrom (2005) y que sostiene que el conocimiento tenga que ser protegido ante la tendencia de su mercantilización.

Durante el texto se irán revisando las aportaciones de Fricker (2007) y Goldman (1999), se revisarán ejemplos concretos como repositorios abiertos (arXiv, CORE, PubMed Central), y se examinarán tensiones entre el acceso libre y la lógica del lucro en el ámbito académico. No se parte de intenciones tecnofóbicas ni tecnofílicas, sino de preguntas éticas y políticas: revisar las mediaciones que hoy determinan qué conocimiento es visible, accesible y legítimo, y abrir caminos para una inteligencia artificial al servicio del bien común.

La emergencia por COVID-19 precipitó de manera drástica la digitalización de la vida académica, las universidades hicieron el tránsito a plataformas virtuales, la producción científica se incrementa exponencialmente, la necesidad de acceder a información confiable se distribuye como cuestión de salud pública. En este sentido, los sistemas de inteligencia artificial asumieron el protagonismo para la difusión del conocimiento bibliográfico, organizando miles de artículos, generando resúmenes automatizados, conectando investigadores al mismo tiempo. Pero también se evidencia la fragilidad del acceso a intensas conexiones digitales e idiomas vernáculos: era una realidad evidente que las personas que no contaban con conectividad, competencias digitales o dominio del inglés se encontraban excluidas. Esta experiencia global — aún no tan distante— reanima la revisión del propio conocimiento en el marco de infraestructuras digitales en clave de una ética del cuidado, la justicia, la inclusión,

reforzamiento democrático y la inclusión digital contemporánea. Se aborda la forma en que los sistemas automatizados de información o los modelos de lenguaje como ChatGPT suponen un nuevo aparato técnico que hace vibrar las maneras de acceso, de producción y de validación del conocimiento a partir de conceptos como la responsabilidad epistémica; la injusticia algorítmica o la equidad cognitiva. Al hilo de este análisis surgen advertencias respecto a las tensiones entre el conocimiento como un bien común y su progresiva manera mercantil de transformarse por el propio sistema de conocimiento, que da visibilidad a ciertas formas de resistencia como los repositorios abiertos, el movimiento por el acceso libre o las infraestructuras de saber colaborativo. El capítulo concluye con la propuesta de una ética digital del conocimiento a partir de la construcción basada en los principios de transparencia algorítmica, pluralidad epistémica y justicia cognitiva, y entendida como una condición necesaria para el reforzamiento democrático y la inclusión en los entornos digitales contemporáneos.

En este sentido, el capítulo propone articular la noción de responsabilidad epistémica con los debates contemporáneos sobre inteligencia artificial y bienes comunes del conocimiento, argumentando que una ética digital del saber requiere no solo transparencia algorítmica, sino también estructuras institucionales orientadas a la justicia cognitiva y al acceso abierto.

## **La responsabilidad epistémica como concepto clave**

Para precisar el concepto de responsabilidad epistémica, resulta útil acudir a la epistemología de las virtudes. Desde esta tradición, conocer no consiste únicamente en poseer creencias verdaderas y justificadas, sino en ejercer disposiciones intelectuales responsables al buscar, evaluar y transmitir conocimiento. Lorraine Code (1987) subraya que el conocimiento es una práctica situada que implica obligaciones morales e intelectuales por parte de los agentes cognitivos. En una línea afín, Linda Zagzebski (1996) sostiene que virtudes epistémicas como la honestidad intelectual, la apertura crítica y la diligencia cognitiva orientan a los sujetos hacia la verdad y la comprensión. En el contexto contemporáneo de mediación algorítmica del saber, estas virtudes no conciernen solo a individuos, sino también a instituciones y a infraestructuras tecnológicas que participan en la producción y circulación del conocimiento, lo que permite pensar la responsabilidad epistémica como una práctica distribuida.

Cuando hablamos de conocimiento, normalmente creemos que solo nos referimos a verdades universales, verdades factuales, proposiciones fielmente fundamentadas, argumentos de razón. Pero muy pocas veces nos paramos a pensar cómo llegan a ser —o cómo llegan a no ser— esos conocimientos, cuáles son las condiciones sociales, políticas y tecnológicas que explicarían cómo saberse compartido,

recordado o validado. De hecho, en escenarios donde la información circula rápidamente, pero sin la seriedad de ser legitimada, la noción de responsabilidad epistémica se vuelve incluso crucial: ¿quién responde cuando el saber se deforma, se hace un saber sesgado, se viraliza de manera desigual?

Con relación a la preocupación estructural ya mostrada por Bourdieu (1975), el campo científico no es un espacio neutro, sino un espacio de luchas simbólicas en el que el capital académico, el prestigio institucional y las reglas de validación determinan qué voces son escuchadas y cuáles son descalificadas. Esta lógica de poder, que operaba en revistas, congresos y comités editoriales, hoy se amplifica mediante algoritmos que regulan el flujo digital de los saberes. Comprender la injusticia epistémica asociada a la IA requiere articular lo micro (decisiones individuales) con lo macro (estructuras de legitimación) y mostrar cómo ciertas plataformas consolidan jerarquías ya existentes.

Desde la epistemología social, la pregunta ha cobrado gran importancia en las últimas décadas. Así, por ejemplo, Fricker (2007) propone el concepto de injusticia epistémica, que describe las situaciones donde alguien es desautorizado como sujeto que sabe o conoce a partir de "los prejuicios sociales". La injusticia epistémica no es simple carencia de información, sino que es una forma de exclusión simbólica: se da lugar a situaciones donde una persona es desacreditada de forma sistemática —en tanto que esa persona es una mujer,

pertenece a una etnia determinada, es de una clase baja determinada o emite su enunciado desde un lugar de enunciación/categoría simbólica determinado— cuya conclusión es negar a esa persona la evidencia de juicios, de testimonios, considerados como válidos o que otorgan sustancia a la imposición de ser la fuente del saber que se quiere colocar entre la gente.

En el mundo digital, la injusticia toma otras formas. No consiste solamente en que un individuo discrimine a otro en una conversación, sino que son sistemas algorítmicos los que, al estar entrenados con datos históricos sesgados, reproducen o incluso aumentan patrones de invisibilización. Así como ha advertido Slussareff (2022) muchos de los algoritmos que están al frente de nuestros rankings, sugerencias y filtros están desarrollados sin un sentido crítico hacia sus implicaciones sociales. Las voces, este tipo de voces, marginadas tienden a quedar fuera de la red de la digital visibility, lo que nos hace llegar a una nueva forma de exclusión: la injusticia algorítmica.

En una época en la que la cantidad de información a la que podemos acceder es abrumadoramente superior a las capacidades humanas de la cognición, decir que se ha vuelto necesario pensar el conocimiento como una práctica situada, mediada y relacional, no solo como cuestión de contenido. La responsabilidad epistémica se convierte, en este orden de ideas, en una ética que se vuelve algo más que un hecho

individual porque no se refiere únicamente a “estar bien informado”, sino que permite reflexionar sobre las condiciones sociales, culturales y tecnológicas que definen quién puede conocer, qué se considera conocimiento legítimo y cuáles son las bases para que este último se valide. En este panorama es en el que se hace necesario pensar las implicaciones filosóficas de esa responsabilidad epistémica, es decir, tratando a la vez los mecanismos de exclusión, así como los retos actuales que la inteligencia artificial plantea para la justicia epistémica.

### *Entre Fricker y Medina: injusticia epistémica estructural*

Fricker (2007) introduce dos formas de injusticia epistémica: la injusticia testimonial (que ocurre cuando se niega la creencia a las personas por prejuicios) y la injusticia hermenéutica (que tiene lugar cuando a una persona o grupo les faltan recursos conceptuales para dar significado a su experiencia). Ambos tipos de injusticia tienen relevancia en la actualidad, pero Zobel (2019) es quien enriquece esta propuesta con la idea de que la injusticia epistémica no es sólo un fenómeno que ocurre entre personas, sino también una estructura histórica y sistémica que reproduce tanto silenciamientos como desigualdades.

Zobel (2019) trae a la escena la idea de resistencia epistémica, que alude a la capacidad crítica que tienen las comunidades de cuestionar los trazos de los marcos opresivos

(es decir, hay oportunidad de que generen nuevos lenguajes, categorías y prácticas de conocimiento). Esta idea es un elemento básico si lo que se quiere es avanzar hacia una IA éticamente responsable, pues no basta con “corregir” algoritmos (que son más “justos”) sino que, además, es necesario transformar las lógicas sociales y políticas mediante las cuales se determinan qué saberes son válidos.

En este sentido, la responsabilidad epistémica no es simplemente individual (del programador, del investigador o del usuario), sino también colectiva y estructural. Esto implica diseñar entornos tecnológicos que no reproduzcan jerarquías de credibilidad, sino que promuevan (activamente) la pluralidad de voces y experiencias.

### *Goldman, Latour y la fiabilidad distribuida*

Desde un punto de vista diferente, el enfoque veritista de la epistemología social propuesto por Goldman sostiene que las creencias que circulan en una comunidad deben tender hacia la verdad, lo que requiere instituciones y prácticas epistémicas confiables. Para Goldman, el mecanismo social (revistas, expertos, revisiones) tiene que estar construido de una forma que la fiabilidad epistémica tienda a maximizarse. Eso entraña una forma de responsabilidad: cada agente que contribuye en la cadena del conocimiento lo hace en integridad, de la mano de la transparencia, de la honradez con respecto a la verdad.



Es un modelo que resulta interesante, pero quizás no suficiente en situaciones donde los agentes ya no sean exclusivamente humanos, como sugiere Latour (2005) desde la teoría del actor-red, el conocimiento es el resultado de interacciones entre humanos y no-humanos (tecnologías, archivos, objetos, programas). En este sentido, un motor de búsqueda, un algoritmo de recomendación o un modelo de lenguaje como ChatGPT no son meras herramientas, sino que se constituyen en agentes que participan activamente en la construcción del saber.

Si suponemos que esta es la situación, correspondería entonces pensar la responsabilidad epistémica como algo distribuido, pero no únicamente del autor que redacta o del lector que interpreta, sino también del sistema técnico que media entre ambos; y, en consecuencia, eso supone nuevas maneras de rendir cuentas, nuevas maneras de certificar la transparencia algorítmica, nuevas maneras de establecer una regulación ética que tenga en consideración las posibilidades de impacto de esos actores no humanos.

## **Inteligencia Artificial y nuevas formas de acceso al conocimiento**

El desarrollo de tecnologías de inteligencia artificial (IA) en la vida cotidiana ha transformado profundamente la forma en la que se produce, organiza y accede al saber. Desde sistemas de recomendación de la información ajustados a los

perfiles de usuarios en los motores de búsqueda hasta asistencias inteligentes como ChatGPT, la mediación tecnológica no es un canal neutral, ha pasado a formar parte del saber.

Esta mediación tecnológica no es inocente, como ya decía Zuboff (2020), estamos bajo una lógica de capitalismo de vigilancia en la que nuestros comportamientos, búsquedas e intereses intelectuales son de forma continua recogidos, procesados y monetizados por las grandes corporaciones. El conocimiento ya no es un fin en sí mismo, sino un insumo más para una economía de datos. Las plataformas que prometen un uso de la información que priorice el acceso a la misma también se benefician por cada clic, cada pregunta formulada, cada tema explorado. Por lo que pensar la ética de la IA en la difusión del saber exige también un ejercicio de crítica de la lógica económica que la sostiene.

Las tecnologías de inteligencia artificial no solo han transformado nuestras prácticas de acceso al conocimiento, sino también en nuestras maneras de buscar, de interpretar y de reproducir estos conocimientos. En esta línea, el hecho mismo del acceso ya no depende sólo de bibliotecas o de redes de investigación, sino que se encuentra determinado por unas interfaces algorítmicas que filtran la realidad a partir de criterios oscuros. Ciertamente, esta transformación promete democratizar el saber, pero, al mismo tiempo, también se plantean preguntas éticamente urgentes sobre la mediación

tecnológica y el papel que tienen estas plataformas en el proceso de construcción del pensamiento público. En las subsecciones que siguen, se expondrán tanto las amenazas como las posibilidades que la IA llega a suponer para el ecosistema cognitivo mundial.

### *Plataformas basadas en IA y circulación del saber*

El acceso a la literatura científica ha llegado a millones de personas en todo el mundo gracias a herramientas como Google Scholar, Semantic Scholar, Lens.org o Dimensions.ai. Éstas permiten localizar obras de un/a autor/a, por tema, por citas, por afiliaciones o por palabras clave; su disponibilidad es casi instantánea. Pero su disponibilidad es el resultado de algoritmos que enfatizan unos resultados específicos, que ordenan la relevancia por criterios cuantitativos, como el número de citas, o que filtran los contenidos por medio de los idiomas, del tipo de fuente, o de la visibilidad institucional.

Esta lógica de priorización automatizada tiene implicaciones éticas de gran escala. Como advierte Zobel (2019), los algoritmos no son neutros ni objetivos: se entrenan con datos que reflejan desigualdades del mundo real y, por tanto, tienden a reproducirlas. Esto puede traducirse en sesgos raciales, de género o geográficos que inciden en el tipo de conocimiento al que accedemos y en la visibilidad de ciertas comunidades epistémicas.

No es que no existan estas plataformas —que han democratizado en parte el acceso; lo verdaderamente problemático es su opacidad: los usuarios desconocen hasta qué punto se les ofrece un artículo sí y otro no y no tienen ningún control sobre los filtros que determinan su búsqueda. En ese sentido, se compromete el principio de agencia epistémica: las personas pierden su condición de sujetos activos en el conocimiento y se convierten en receptores pasivos de una curaduría automatizada.

*Oportunidades y riesgos: el rol de los algoritmos en la mediación cognitiva*

Desde la ética de la tecnología se ha insistido en que los sistemas digitales reconfiguran prácticas de juicio y, con ello, las condiciones de responsabilidad. Vallor (2016) propone pensar las tecnologías emergentes desde la formación de virtudes tecnomorales —incluidas virtudes intelectuales— capaces de sostener agencia crítica en entornos sociotécnicos complejos. Aplicado a la difusión del conocimiento, esto sugiere que no basta con exigir precisión factual: también es necesario interrogar el diseño de las plataformas, su transparencia y los incentivos económicos que orientan su mediación del saber.

El avance de la inteligencia artificial también ha brindado posibilidades interminables, herramientas como Elicit, Scite o Consensus hacen posible formular preguntas en

lenguaje natural y buscar artículos científicos, comparar resultados de estudios e incluso verificar la validez de tales afirmaciones a partir de la literatura revisada por pares. Estas aplicaciones son especialmente beneficiosas para estudiantes, investigadores en la etapa inicial o lectores no especializados que deseen acceder a información científica, pero que no se ven en condiciones de dominar los lenguajes técnicos de cada especialidad. No obstante, al lado de estas posibilidades, también afloran nuevos peligros. Uno de ellos es el empobrecimiento del juicio crítico. Si los sistemas responden con convencimiento ante cualquier pregunta, el riesgo de aceptar sus aportaciones se multiplica, incluso si son inadecuadas o sesgadas. Esto resulta ser muy grave en los modelos generativos de ChatGPT, por ejemplo, que no son capaces de discriminar entre conocimiento verificado e información plausible, y que, sin embargo, logran generar afirmaciones falsas con un alto grado de fluidez y con coherencia formal.

Para añadir, gran parte de estos sistemas han sido entrenados con datos obtenidos a partir de fuentes cerradas o con representaciones desiguales, lo que puede afectar directamente a la cobertura temática y a la diversidad epistémica. Tal y como advierten Gebru et al. (2021) en su propuesta de “fichas técnicas para datasets” (Datasheets for Datasets), el hecho de que no tengamos claro el origen y la composición de los datos que sirven para entrenar modelos de

IA es un reto, una amenaza directa a la transparencia y justicia informativa.

Pero también hay que notar la creciente dependencia de estos sistemas en contextos educativos, institucionales y científicos. Cuando las universidades empiezan a recurrir a IA para revisar un texto, para recomendar bibliografía, o incluso para redactar un informe, la pregunta que surge es: ¿qué criterios éticos y pedagógicos deberían implementarse?, es decir, ¿qué significa aprender o investigar, si una herramienta puede hacer el trabajo por nosotros? ¿Estamos quizás ignorando la responsabilidad de delegar a sistemas que no pueden entender el valor humano del conocimiento?

En este contexto, la IA es una herramienta de expansión cognitiva y también una fuerza de estandarización y automatización del saber. Se torna un desafío encontrar ese punto medio entre la accesibilidad y la crítica, entre la automatización y la reflexión, entre la velocidad y la profundidad. Como sugiere Zuboff (2020), estamos actualmente en una economía del control, donde los datos de la persona y los patrones de conducta son la nueva moneda. En esta situación no escapan a ello ni nuestras búsquedas intelectuales, las cuales son una materia prima para algoritmos cuya intención no es la verdad sino la rentabilidad.

## **El conocimiento entre el bien común y la mercantilización**

Habitualmente afirman los discursos institucionales que el conocimiento es un derecho humano, una palanca del desarrollo sostenible, un motor de la democracia. No obstante, estas afirmaciones contraponen de forma alarmante las prácticas concretas de distribución de saberes, predominantemente en el ámbito académico. El saber, lejos de ser de acceso libre y gratuito, permanece encerrado tras muros de pago en el caso de no ser controlado por las grandes editoriales científicas y redistribuido a partir de lógicas comerciales muy distantes del bien común.

Suber (2003), es el principal abanderado del movimiento del acceso abierto, afirma que el conocimiento financiado con fondos públicos tiene un deber ético irrenunciable: todo conocimiento financiado con fondos públicos debe devolver a la sociedad la cual lo ha hecho posible. Para él, el acceso abierto es una cuestión de necesidad moral y no simplemente una cuestión de facilitar la visibilidad científica del mismo, sino una exigencia moral en consonancia con la equidad y mostrar la responsabilidad pública, considerando que la negativa del acceso a la ciencia misma —sobre todo en cuestiones de emergencia social como son la salud, el medio ambiente, o los derechos humanos— solo perpetúa la desigualdad y debilita el ejercicio informado de la ciudadanía.

No obstante, el hecho de que ha sido entendido históricamente el conocimiento como fuente de la emancipación humana, en las últimas décadas se ha afianzado una tendencia que resulta preocupante: la conversión del conocimiento en mercancía bajo lógicas de acumulación, exclusión y rentabilidad. El mercado de la educación científica, la propiedad intelectual restrictiva, las plataformas comerciales han coartado la posibilidad de acceder a la información científica incluso en el seno de la academia. Esta sección explicita el conflicto que se establece entre el ideal del conocimiento como bien común y su apropiación por parte de estructuras corporativas, subrayando las barreras al acceso, las resistencias que se van estableciendo y los modelos éticos alternativos.

### *Modelos de negocio y barreras de acceso*

El modelo mayoritario de la publicación científica ha sido duramente criticado por su estructura paradójica: los investigadores, la mayor parte de los cuales son financiados con recursos públicos, producen posteriormente artículos que son entregados gratuitamente a las casas editoriales privadas, las cuales cobran además altos precios tanto a los centros de enseñanza universitaria por el acceso a su contenido como a los propios autores por el hecho de poder publicar sus resultados de investigación. En este modelo, el conocimiento se convierte en una mercancía objeto de un valor de cambio y se aleja de ser considerado un bien común.



El caso de Elsevier, una de las editoriales académicas más poderosas del mundo, ha sido especialmente ilustrativo, dado que esta compañía, que opera con márgenes de beneficios superiores al 30% anual y goza de una posición casi monopólica, ha sido objeto de críticas sin precedentes y ha generado movimientos como #boycottElsevier impulsados por los académicos que se niegan a publicar, revisar o editar para revistas que no permiten el libre acceso al conocimiento que generan. La crítica no es nueva, pero sí se ha intensificado, especialmente en los últimos años al compás de la expansión de las redes digitales, las cuales han sido capaces de poner de manifiesto que la única barrera técnica para el acceso al conocimiento ya ha sido superada. No obstante, lo que queda son barreras políticas, jurídicas y económicas.

Un ejemplo paradigmático de resistencia ha sido Sci-Hub, una plataforma que fue creada por las investigadoras de origen kazajo Elbakyan y Bozkurt (2021) para dar acceso gratuito a millones de artículos y diversas publicaciones académicas. Aunque está siendo denunciada por las editoriales de revistas científicas por haber violado derechos de autor y derechos de propiedad intelectual, es defendida por amplios colectivos como un perfecto instrumento de justicia cognitiva, sobre todo en países del Sur Global. En su carta abierta, dirigida a la comunidad científica, la propia Elbakyan defiende que el conocimiento científico se crea con fondos públicos; en consecuencia, debe pertenecer a la humanidad, no a unas corporaciones privadas.

### *Acceso abierto como principio ético*

Frente a este proceso de mercantilización del conocimiento, en contraposición, ha emergido el movimiento de promoción de acceso abierto (open access), una alternativa que busca posicionarse desde alguna ética, política y epistemológicamente hablando. Esto es, los resultados de la investigación científica deben estar accesibles sin restricciones económicas, técnicas o legales. Y de diversas iniciativas, Plan S, promovido por una coalición de agencias financiadoras europeas que establece que la investigación financiada con fondos públicos debe publicarse en open access a partir de 2021.

Autores como Suber (2003) han sido voces significativas a la hora de defender el open access como un derecho epistémico, siendo que Suber (2003) sostiene que el open access no es solo una política científica, sino que es una forma de intervenir y corregir desigualdades históricas en la producción y circulación del saber, donde el conocimiento debe orientarse desde el poder adquisitivo hacia el beneficio común y la cooperación internacional. Se definen diversos modelos de open access: repositorios institucionales, revistas sin costes de publicación ni suscripción (modelo diamond OA), pero también plataformas colaborativas como Open Access Button y Unpaywall que permiten localizar versiones libres de los artículos académicos, desafiando así las restricciones que imponen editoriales tradicionales.

Sin embargo, el movimiento de acceso abierto tiene sus tensiones. Algunos modelos como el gold OA transfieren el coste al autor, exigiendo altas tarifas por publicar. Se crea una nueva exclusión: los investigadores sin financiamiento quedan fuera de la visibilidad académica, reproduciendo la misma lógica de exclusión elitista que se pretendía combatir; por eso, no se puede limitar el debate ético sencillamente al acceso sino incluir las condiciones de participación, producción y reconocimiento del conocimiento. En este sentido, el conocimiento como bien común se plantea no solo como un ideal normativo sino como una posición política concreta respecto de las dinámicas de apropiación corporativa del saber. Como dicen Hess y Ostrom (2005), los bienes comunes cognitivos urgen no solo de apertura técnica, sino de gobernanza compartida, reglas de éticas claras y comunidades responsables de su cuidado. La IA ha permitido organizar y recuperar el conocimiento, pero también puede ser un mecanismo de cierre si sus bases de datos, algoritmos de búsqueda o resultados generativos dependen solamente de fuentes cerradas y para ello una inteligencia artificial ética debe incorporar desde su diseño principios de acceso abierto, diversidad epistémica y responsabilidad social.

### **Repositorios abiertos, saber compartido y equidad epistémica**

Si la idea de la mercantilización del conocimiento puede llegar a ser una amenaza para la equidad epistémica,

los repositorios abiertos han emergido como espacios de resistencia y esperanza. Estas plataformas, por un lado, permiten el libre acceso a contenidos científicos y, por otro lado, encarnan una ética del conocimiento fundada en la colaboración, la inclusión y la reciprocidad. En un mundo profundamente desigual en términos cognitivos, los repositorios abiertos representan un intento significativo de democratización del saber.

Ennis (2007) refuerza esta postura cuando expresa que “el principio del acceso al conocimiento es un imperativo ético, no sólo un derecho académico.” En su análisis, limitar el acceso a la investigación genera un empobrecimiento del diálogo público, impide el desarrollo de la innovación y quiebra la relación de la ciencia con la sociedad. Desde esta visión, los repositorios abiertos no son sólo soluciones técnicas que permiten su funcionamiento, sino manifestaciones de una ética del compartir, en la que el saber se entiende como un bien relacional, un bien que pasa entre personas, que tiende a circular, a dialogar y a contribuir a la transformación social.

A raíz del cerco comercial que coarta el acceso a la producción científica, también han surgido movimientos que reclaman el saber como un bien que ha devenir común, colectivo y sustentado por los principios de la equidad y la colaboración. Los repositorios abiertos no persiguen solamente la eliminación de los muros de pago, sino que

también desafían las jerarquías epistémicas dominantes, proponiendo una lógica de la distribución del saber más equitativa e inclusiva. En este apartado vamos a exponer algunas de estas iniciativas, poniendo de manifiesto su impacto, sus limitaciones y su potencial a la hora de convertirse en herramientas de transformación estructural.

### *Plataformas como arXiv, CORE, PubMed Central, DOAJ*

El caso de arXiv sirve de modelo, fundado en 1991 por Paul Ginsparg, se convertiría en uno de los primeros repositorios abiertos en Internet, dando la oportunidad a un conjunto muy numeroso de físicos y matemáticos de compartir antes de la publicación la mecánica de sus trabajos. Hoy, arXiv contiene más de dos millones de artículos en física, matemática, computación y otras áreas, y se ha convertido en un espacio de referencia a nivel mundial para investigadores que privilegian la rapidez, la transparencia y el acceso abierto.

Al lado de arXiv han ido apareciendo otras plataformas que han dado sustento a un ecosistema de repositorios en línea. Con sede en el Reino Unido, CORE (Connecting Repositories) recolecta millones de artículos de miles de repositorios institucionales de todo el mundo, permitiendo la búsqueda, el descubrimiento y el acceso sin barreras al contenido académico. PubMed Central, también gestionado por la Biblioteca Nacional de Medicina de EE.UU., se configura como un repositorio garantizado de literatura

biomédica de acceso libre. Más allá de la cantidad, lo más interesante de estos espacios es su filosofía inspiradora: el conocimiento como recurso colectivo no como producto de mercado.

El Directory of Open Access Journals (DOAJ) opera como un catálogo internacional que incluye revistas científicas de acceso abierto y que demanda criterios de calidad en la edición. Este proyecto no solo promueve acceso sino también confianza, robusteciendo la fe en modelos no comerciales de publicación. Estas plataformas, junto con otras herramientas como Unpaywall y Open Access Button, crean un entorno en expansión que disputa la posición dominante de los grandes grupos editoriales.

### *El conocimiento como patrimonio colectivo*

Si bien el conocimiento como patrimonio colectivo implica su consideración como un bien común, donde el valor no reside en su escasez o apropiación sino en su circulación, enriquecimiento y reapropiación por las comunidades, e incluso tal como defendían las autoras Hess y Ostrom (2005), el conocimiento compartido no requiere solo acceso abierto, sino gobernanza colaborativa, responsabilidad compartida y mecanismos éticos para su cuidado.

Los repositorios abiertos no siguen la lógica de depender de una única empresa, sino que articulan redes

institucionales y comunitarias que sostienen su funcionamiento. No son concebidos como escaparates de publicaciones, sino como infraestructuras públicas de conocimiento, equivalentes a las bibliotecas, pero adaptadas a la digitalización. Ahora bien, estos repositorios no están exentos de retos. Uno de ellos es la brecha tecnológica: en muchas universidades del Sur Global no hay servidores, conectividad o personal que mantenga un repositorio. Otro problema lo constituye el lenguaje hegemónico: una parte sustantiva del contenido —no solo de artículos, sino también de repositorios de datos o de comunicaciones científicas— está en inglés, lo cual produce invisibilización de saberes locales, experiencias regionales y formas de epistemologías no eurocentradas (Ramírez-Castañeda, 2020).

Además, persisten incertidumbres respecto al reconocimiento académico de los documentos depositados en repositorios. Si bien muchas revistas aceptan el “preprint” como una versión válida, en algunas disciplinas o instituciones estos documentos no poseen el mismo prestigio que los artículos en revistas indexadas. Esta situación reactiva la tensión entre una noción de acceso libre y otra de capital académico, donde el valor simbólico del conocimiento sigue mediado por estructuras jerárquicas.

Así las cosas, el avance hacia una equidad epistémica real y efectiva no sólo abre el acceso, sino que implica asimismo revisar los criterios de legitimidad, de autoría o de

validación del saber. Como ha señalado de Sousa Santos (2009), estamos sometidos a un “monocultivo del saber científico” que impide visibilizar otras formas de saber. Podría ser una buena forma para ampliar ese campo pero que incluye también saberes indígenas, comunitarios, feministas y otras formas de producción de conocimientos que quedan, al mismo tiempo, al margen de los marcos dominantes.

La inteligencia artificial puede contribuir a la expansión del acceso si se utiliza de manera responsable: puede funcionar como herramienta inclusiva si se alimenta de repositorios abiertos y aprende a valorar la diversidad epistémica. Pero si se entrena exclusivamente con fuentes comerciales o cerradas, tenderá a reproducir las exclusiones ya conocidas. Por eso, una IA ética debería nutrirse de bienes comunes cognitivos y trabajar con datos abiertos, auditables y representativos; solo así podrá contribuir a una justicia epistémica que no sea retórica, sino estructural.

## **Conclusiones y proyecciones éticas**

El saber, en todas sus versiones o acepciones, es una de las más sometidas a la mirada del ser humano. A través del saber nos entendemos, nos interpelamos, nos proyectamos. Sin embargo, en la actualidad el propio saber, el que hemos ido acumulando, ordenado y difundido, empieza a estar condicionado por lógicas algorítmicas, intereses comerciales y desigualdades sistémicas que fijan la condición de acceso al



saber, de hacer uso de este en cualquier sentido y de legitimar este uso. En el presente capítulo hemos seguido el rastro de algunas de las principales tensiones que delimitan la divulgación del saber en tiempos de inteligencia artificial. Hemos constatado cómo las herramientas tecnológicas no son jamás neutrales, sino que suceden nuevas formas de exclusión o formas de reconocimiento; cómo los sistemas basados en IA configuran lo que leemos, en lo que creemos y quiénes somos quienes escuchamos; y, para terminar, cómo la mercantilización del saber acaba convirtiéndolo – el saber, bien común por C – en privilegio de unos pocos.

Y en esta misma dirección, la Norma de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura sobre Ética de la Inteligencia Artificial (UNESCO, 2022) establece que los sistemas automatizados deben ser diseñados y regulados de acuerdo a principios de inclusión, equidad, responsabilidad y participación. Este marco normativo establece que el desarrollo de la inteligencia artificial no se debe orientar solamente a la eficiencia o la innovación técnica, sino también a la promoción de los derechos humanos, de la diversidad cultural y la equidad cognitiva. Así, defender una ética del conocimiento en la era digital es también conformarse con una visión del mundo que asume el saber cómo un derecho universal y no como un privilegio disperso.

Por lo tanto, abogamos por considerar la responsabilidad epistémica como un principio ético básico. Pero no como la responsabilidad individual, aislada o abstracta, sino como responsabilidad colectiva, estructural y contextual. No es suficiente pedir precisión a las plataformas tecnológicas, tenemos que exigirles que operen a partir de la asunción de los principios éticos de equidad, transparencia y pluralismo. Como comunidad académica, como ciudadanía crítica, como humanidad interconectada nos hacemos responsables de recuperar el conocimiento como bien común, como patrimonio colectivo y como derecho inalienable.

Las reflexiones que se exponen aquí nos empujan a revisar detenidamente las condiciones bajo las cuales hoy circula el conocimiento. La tecnología, y en concreto la inteligencia artificial, no puede ser considerada como un mero canal de información, sino que debe concebirse como un lugar de litigio político, ético y de conocimiento. Por eso, no basta con mejorar los sistemas técnicos, sino que resulta necesario imaginar y construir una cultura del conocimiento más justa, plural y transparente. A lo largo de lo que sigue se trazan una serie de principios orientadores para avanzar hacia una ética digital del conocimiento que responde a los desafíos existentes y que abre horizontes más igualitarios hacia el futuro.

## *Propuesta de principios para una ética digital del conocimiento*

Para concluir, propongo algunos principios que podrían guiar una ética de la difusión del conocimiento en el contexto de la digitalización con predominio de determinada IA. No son recetas, sino horizontes posibles. Transparencia algorítmica: los sistemas de recomendación, búsqueda o síntesis deben poder ser auditados; sus criterios de priorización tienen que ser accesibles y promovidos por los propios usuarios. La opacidad no puede ser el precio de la eficiencia. Pluralidad epistémica: el conocimiento no es uniforme ni unilingüe; las plataformas tienen que incorporar y visibilizar saberes diversos: científicos, indígenas, comunitarios, feministas, alternativos. La diversidad es condición de justicia cognitiva. Acceso abierto como norma, no como excepción: la totalidad de las investigaciones financiadas con dinero público tiene que ser accesible al público, lo cual incluye artículos, datos, materiales docentes y software. La ciencia no puede estar sometida a la lógica de la privatización.

Responsabilidad compartida: Durante un largo tiempo, no solo los programadores o los editores han tenido que ser responsables de la circulación en el campo del conocimiento. Universidades, gobiernos, plataformas e individuos son partes responsables del cuidado del ecosistema cognitivo que disfrutamos. Educación crítica en

tecnología del saber: La alfabetización informacional debe incluir una comprensión de la ética de los sistemas de IA. No basta con saber utilizarlos, sino que también hay que interrogar y regular y mejorar. Estos principios no se agotan en lo técnico. Son, al fin y al cabo, propuestas políticas. Porque cualquier forma de organizar el conocimiento, como la forma de decidir qué se dice, quién lo dice, cómo se dice y para quién se dice, es también, una forma de organizar el mundo.

### *El papel ambivalente de la IA: ¿aliada o amenaza?*

La inteligencia artificial posee un enorme potencial para democratizar el acceso al conocimiento. Entre sus capacidades más destacadas se encuentran la superación de barreras lingüísticas, la simplificación de lecturas complejas, la organización de información dispersa y la recomendación de contenidos relevantes según intereses específicos. Estas funciones pueden abrir nuevas oportunidades para comunidades históricamente excluidas del circuito científico y académico.

Sin embargo, este mismo potencial encierra riesgos considerables. La inteligencia artificial también puede transformarse en una herramienta de homogeneización cognitiva, de vigilancia constante y de concentración del poder epistémico en manos de unos pocos actores tecnológicos.

La ambivalencia de la inteligencia artificial es innegable: sus efectos dependerán, en última instancia, del modo en que sea diseñada, de los sujetos para quienes esté concebida y de los valores que orienten su funcionamiento. Una IA entrenada con datos diversos, desarrollada bajo principios éticos y regulada desde marcos de justicia puede contribuir significativamente a la expansión de nuestras capacidades colectivas. Por el contrario, una inteligencia artificial opaca, mercantil y excluyente no hará sino reforzar las desigualdades que decimos querer erradicar.

En el contexto actual —marcado por la posverdad, la desinformación y la sobrecarga informativa— defender el conocimiento como bien común constituye no solo un acto de responsabilidad intelectual, sino también un compromiso con la justicia social. Esta defensa exige algo más que garantizar el acceso: implica asegurar la participación activa en la producción del saber y revisar críticamente las condiciones que legitiman su validación y circulación.

Necesitamos, más que nunca, recuperar el sentido del saber cómo una práctica ética y colectiva. Esto exige cultivar una auténtica ecología del conocimiento, en la cual la inteligencia artificial no sustituya al pensamiento crítico, sino que lo complemente, lo acompañe y lo potencie. Ese es el verdadero reto. Y, posiblemente, nuestra tarea más urgente.

En última instancia, la ética digital del conocimiento debe ir más allá de consideraciones meramente técnicas o académicas: debe asumirse como una tarea pedagógica y cultural. Educar ciudadanos capaces de interactuar críticamente con la información, cuestionar sus fuentes, exigir transparencia y participar activamente en el diseño de las tecnologías que los afectan, se erige hoy como una de las obligaciones éticas más apremiantes. La educación, en todos sus niveles, debe formar no solo competencias digitales, sino también una conciencia crítica sobre los saberes que se producen, se legitiman y se consumen en red.

## Referencias

- Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, 14(6).  
<https://doi.org/10.1177/053901847501400602>
- Code, L. (1987). *Epistemic responsibility*. University Press of New England.
- de Sousa Santos, B. (2009). *Una epistemología del Sur: La reinención del conocimiento y la emancipación social*. Siglo XXI / CLACSO.
- Elbakyan, A., & Bozkurt, A. (2021). A critical conversation with Alexandra Elbakyan: Is she the pirate queen, Robin Hood, a scholarly activist, or a butterfly flapping its wings? *Asian Journal of Distance Education*, 16(1).

- Ennis, L. A. (2007). The access principle: The case for open access to research and scholarship. *Journal of the American Society for Information Science and Technology*, 58(9).  
<https://doi.org/10.1002/asi.20594>
- Foucault, M. (1972). *The archaeology of knowledge*. Pantheon.
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12). <https://doi.org/10.1145/3458723>
- Goldman, A. I. (1999). *Knowledge in a social world*. Oxford University Press.  
<https://doi.org/10.1093/0198238207.001.0001>
- Hess, C., & Ostrom, E. (2005). A framework for analyzing the knowledge commons. En *Understanding knowledge as a commons: From theory to practice*.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.  
<https://doi.org/10.1093/oso/9780199256044.001.0001>
- Ramírez-Castañeda, V. (2020). Disadvantages in preparing and publishing scientific papers caused by the

dominance of the English language in science: The case of Colombian researchers in biological sciences. *PLoS ONE*, 15(9), Article e0238372. <https://doi.org/10.1371/journal.pone.0238372>

Slussareff, M. (2022). [Reseña del libro *Weapons of math destruction: How big data increases inequality and threatens democracy*, por C. O'Neil]. *CyberOrient*, 16(1). <https://doi.org/10.1002/cyo2.26>

Suber, P. (2003). Scholarly communication: Removing barriers to research: An introduction to open access for librarians. *College & Research Libraries News*, 64(2). <https://doi.org/10.5860/crln.64.2.92>

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, 11(2). <https://doi.org/10.1007/s10676-009-9187-9>

UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780190498511.001.0001>

Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

Zobel, G. (2019). [Reseña del libro *Algorithms of oppression: How search engines reinforce racism*, por S. U.



Noble]. *Communication Design Quarterly*, 7(3).

<https://doi.org/10.1145/3321388.3321392>

Zuboff, S. (2020). The age of surveillance capitalism: The fight for a human future at the new frontier of power. *Yale Law Journal*, 129(5).

<https://doi.org/10.26522/brocked.v29i2.849>

# Ética y conocimiento: el problema de la discriminación de los testimonios y los *deepfakes*

*Ethics and Knowledge: The Problem of Discrimination in Testimonials and Deepfakes*

Felipe Álvarez<sup>85</sup>

## Resumen

En este artículo se pretende abordar los problemas éticos y epistemológicos de los *deepfakes* en relación con la discriminación de testimonios de personas naturales, defendiendo que un uso inadecuado de esta tecnología deriva en un daño moral y epistémico hacia las personas simuladas en estos medios sintéticos. Para dar cuenta de ello, se procede de la siguiente manera: en la primera sección se analiza al

---

<sup>85</sup> Universidad de Chile. Fundación Centro Interdisciplinario en Ética, Política y Economía. Financiamiento: Agencia Nacional de Investigación y Desarrollo (ANID-PFCHA/Doctorado Nacional/Año 2022—21220627). <https://orcid.org/0000-0002-7153-2851>

testimonio como fuente de conocimiento; en la segunda, se analiza los problemas éticos y epistemológicos del testimonio a partir del concepto de injusticia epistémica; en la tercera y última, se analizan los *deepfakes* a partir de lo dicho en las secciones anteriores. Con ello se concluye que los *deepfakes* pueden ser nocivos para las comunidades de conocimiento si su uso suscita un daño moral y epistémico a las personas.

**Palabras clave:** Testimonio; *deepfakes*; injusticia epistémica; ética; conocimiento.

## **Abstract**

This article aims to address the ethical and epistemological problems of deepfakes in relation to the discrimination of testimonies from natural persons, arguing that inappropriate use of this technology results in moral and epistemic harm to the persons simulated in these synthetic media. To account for this, we proceed as follows: in the first section, we analyze testimony as a source of knowledge; in the second, we analyze the ethical and epistemological problems of testimony based on the concept of epistemic injustice; in the third and final section, we analyze deepfakes based on what has been said in the previous sections. This leads to the conclusion that deepfakes can be harmful to knowledge communities if their use causes moral and epistemic harm to the people.

**Keywords:** Testimony; deepfakes; epistemic injustice; ethics; knowledge.

## Introducción

A pesar de que en la actualidad uno tenga el conocimiento al alcance de la mano, el testimonio sigue siendo la forma predilecta de adquirir y transmitir lo que sabemos al resto. Lamentablemente, como toda práctica humana, el intercambio testimonial no se encuentra exento de discriminaciones: en efecto, sabemos que existen prejuicios que merman la difusión de conocimiento y que generan daños en las personas. Sin embargo, poco se sabe todavía de cómo el desarrollo de la inteligencia artificial y su uso puede fomentar ese tipo de problemas. Por este motivo, en este artículo se tratará el vínculo indisoluble entre la ética y el conocimiento a partir del problema de la discriminación de los testimonios, proponiendo que los *deepfakes* potencian los casos de injusticia testimonial al servir como una supuesta evidencia para quienes caen en esa clase de prejuicios.

Para lograr ese objetivo, se procederá de la siguiente manera: en la sección 1 se aborda si el testimonio constituye o no una fuente de conocimiento, resaltando la importancia de este para la vida cotidiana; en la sección 2 se hablará acerca de la discriminación de los testimonios a partir del concepto de injusticia testimonial, un tipo de injusticia epistémica en el que los prejuicios impiden considerar la palabra de otros

como una fuente fiable de conocimiento; finalmente, en la sección 3 se problematiza la injusticia testimonial a partir de los *deepfakes* y las posibles consecuencias epistémicas que de ello se derivan. Se pretende, por tanto, dar cuenta de cómo la discriminación de los testimonios es un problema ético y epistemológico que genera daños honra y credibilidad de las personas cuestionadas.

### **¿El testimonio como fuente de conocimiento?**

Como señaló Elizabeth Anscombe (2008), la mayoría de lo que conocemos lo aprendemos gracias a la palabra de otro. Por ejemplo, sabemos qué remedio tomar gracias al médico que lo receta, qué falta está cometiendo mi empleador contra mí gracias al abogado que me defiende o qué sucedió en la batalla naval de Iquique gracias a los historiadores. Si bien todas estas cosas las podríamos saber por cuenta propia si nos dedicáramos a estudiar medicina, derecho o historia; la finitud de la vida humana y la cantidad de asuntos que debemos atender nos impiden ser experto en todas las materias.

Lo anterior suscita, como bien pudo divisar John Hardwig (1985) una dependencia epistémica hacia los demás, a saber, una relación de asimetría intelectual en la que el propio juicio se suspende en pos de atender a un juicio mejor formado. Si uno es racional, dirá Hardwig, deberá hacer caso a lo dicho por el experto en tanto que este último se ha

formado específicamente para esos asuntos a diferencia de uno. De este modo, establecemos relaciones de confianza basada en esa clase de independencia, considerando el testimonio experto como más fiable para la adquisición y divulgación de conocimiento que el testimonio del lego.

Si dependemos de los demás y el juicio experto es mejor que mi propio juicio de lego, cabe preguntarse: ¿el testimonio experto fiable? ¿debo creer necesariamente lo que me dice el experto? ¿es lícito sospechar de ellos? Históricamente ha habido una disputa abierta desde la modernidad hasta nuestros días sobre este asunto: el debate entre las posturas reduccionistas y antireduccionista del testimonio. Analicemos el siguiente caso para contraponer ambas posturas:

#### Caso 1: Martín y el diagnóstico médico

Martín, un barista de una famosa cafetería, va al médico debido a unas lesiones que le han aparecido espontáneamente en su piel. María, la dermatóloga que le tocó atender el caso, le comenta que, dado la ausencia de otros elementos detonantes de la enfermedad (e.g. estrés, agentes externos como perfumes, entre otros), es altamente probable que la causa sea genética. Ante esto, Martín desconfía pues no conoce a nadie en su familia que tenga la enfermedad, de modo que no cree en el testimonio del dermatólogo.

De acuerdo con una postura reduccionista, Martín habría hecho lo correcto. Los reduccionistas defienden que el testimonio no es una fuente epistémica por cuenta propia, de modo que debe ser evaluada a partir de evidencia no testimonialmente fundada como, por ejemplo, la memoria o la sensopercepción (Adler, 2002; Fricker, 2006; Hume, 2007). Martín, al escuchar el testimonio experto, lo ha analizado a partir de sus recuerdos sobre las condiciones médicas de su familia, levantando sospecha sobre la credibilidad del experto en cuanto a la alternativa ofrecida como más probable. Esto, por supuesto, no es necesariamente conducente al escepticismo respecto del testimonio médico, pero sí muestra que se requiere de mayor evidencia no testimonialmente fundada para que el testimonio experto haga sentido a Martín.

Por otra parte, desde una perspectiva antireduccionista, Martín se habría equivocado. Los antireduccionistas defienden que las personas tenemos un derecho a creer en la palabra de los demás porque es natural comunicar lo verdadero sobre lo falso (Baker & Clarke, 2018; Goldberg & Henderson, 2006; Reid, 1983). Según esto, Martín debió haber confiado en el testimonio experto aunque no le hiciera sentido según otras fuentes epistémicas, en tanto que podemos suponer que los otros son honestos por naturaleza y que la mentira constituye más bien una excepción a la regularidad con la que el ser humano tiende a testificar lo verdadero.

Siendo esto así, caemos en la siguiente disyuntiva: ¿cuál postura es correcta para entender en qué medida el testimonio nos conduce al conocimiento? Al parecer, ninguna, pues no captan adecuadamente la naturaleza del intercambio testimonial. Como bien hace ver Ruth Espinosa:

uno de los más grandes errores de los filósofos modernos [y quienes fundamentan sus propias propuestas contemporáneas con base en ellos] fue considerar el testimonio como una forma de más de evidencia, equiparable a la percepción o la memoria. La posibilidad del testimonio reposa, al menos en parte, en las relaciones de confianza que establecemos con otros en tanto cognoscentes. (2021, p. 37)

Otras posturas en epistemología del testimonio como lo son la *assurance view* (Hinchman, 2005; Moran, 2018; Ross, 1986) y la *Trust theory of testimony* (Faulkner, 2011) también han visto este percance, proponiendo alternativas para la resolución de dicho debate. No obstante, la mención de este debate no es más que un pivote para resaltar un problema: el insulto que suscita al oyente el que su credibilidad sea puesta en duda, cuestión que sucedería desde una perspectiva reduccionista como la primera alternativa al caso 1 y que sería un daño hacia su propia persona. En otras palabras, si Martín dudase del médico, ese médico podría verse ofendido por la falta de credibilidad del lego en la situación de dependencia epistémica.



Ante esto tendremos que hacer la siguiente salvedad: no toda duda respecto de los expertos o de alguna persona particular constituye un daño. Bien podría ser que la dermatóloga del caso 1 simplemente se tome con humor la incredulidad de Martín. Sin embargo, el caso toma tintes oscuros si cambiamos algunas de las variables. Si ahora Martín no cree en la dermatóloga porque, supongamos, posee creencias sexistas y experiencias que refuerzan su idea errónea de que las mujeres no son aptas para carreras difíciles como lo es medicina, ya no hablamos de una simple reducción del testimonio a fuentes no-testimonialmente fundadas, sino que hablamos de una reducción que sucedido con base en prejuicio que permea su experiencia y que podría, a la vez, permear las ideas de una cultura completa (e.g. lo que sucede en medio oriente con la educación universitaria de mujeres, guardando las salvedades correspondientes). De ser este el caso, tendríamos que cambiar de foco para preguntarnos cuáles son las condiciones éticas que hacen posible, en mi trato con el otro, la adquisición y divulgación del conocimiento vía testimonio en vez de simplemente preguntarnos si el testimonio es una fuente de conocimiento. A continuación, hablaremos acerca del problema que suscita, por tanto, la discriminación de la palabra de otros.

## **Discriminar la palabra de otros**

La palabra discriminación puede ser entendida en, al menos, dos sentidos: 1. Como un acto de distinción entre la

información recibida. Por ejemplo, ‘Juan discrimina entre noticias verdaderas y *fake news*, de modo que no cae en engaños en la red’; o 2. Como un acto de exclusión respecto de quien funge como fuente de información por su testimonio. Por ejemplo ‘Juan evita juntarse con venezolanos en su trabajo, pues cree que no saben nada y que solo arruinan su país’. Nos interesará principalmente este segundo sentido, sin embargo, debemos reparar en un punto respecto del primero antes de continuar.

Cuando hablamos de discriminar información, es importante recalcar que no necesariamente eso conlleva un daño hacia la persona como tal y que, en la mayoría de los casos, es algo esperado por parte de la audiencia. No se puede simplemente ser crédulo respecto de la información. Analicemos el siguiente caso para ilustrar el punto:

#### Caso 2: Carla y la cura para la depresión según RRSS

Carla, vio en sus RRSS que la depresión se cura consumiendo bebidas energéticas. Al ver el *reel* mientras hacía un transbordo en el metro, no le dio tiempo para hacer *fact checking* y supuso que era información suficientemente fiable por elementos multimodales propios del video (quien lo decía, supongamos, tenía una bata puesta y sonaba profesional). Al juntarse con su amigo Michael, el cual efectivamente ha sido diagnosticado con depresión por

un profesional, Carla le da el consejo de consumir bebidas energéticas para mejorar su condición. Ante esto, Michael discrimina entre lo dicho por Carla y el resto de las creencias que posee al respecto, descartando su testimonio como uno que contiene creencias falsas.

En este caso, no hablamos de un daño para quien testifica, ni a su credibilidad ni a su honra, sino que hablamos meramente de una un acto que, en cuanto a sus consecuencias prácticas y epistémicas de creer en lo dicho por Carla, no es racional para Michael ni creer en lo que dice su amigo por la información que se tiene de antemano. Carlos podría alegar que no se confía en él y tomarlo como una ofensa. Podría, inclusive, decir que Michael lo considera un mentiroso. Sin embargo, ese no sería el caso: nada de esto dice que Michael le tiene por alguien poco fiable ni que lo considera un mentiroso. Bien que podría considerarlo un mero error producto de las circunstancias en las cuales se adquirió la información falsa. Por lo tanto, este acto de discriminación de información no constituye un daño para el hablante propiamente tal, al menos no uno que surja de la audiencia para el hablante.

Cuando hablamos del segundo sentido de discriminar ya nos encontramos con un problema tanto ético como epistemológico. Una cosa es cuestionar lo que dice alguien, y otra muy distinta es cuestionar *a ese alguien*. En el caso

anterior, no se mencionó que Michael dudaba de Carla, sino que se dudaba de lo dicho por Carla porque se tiene información que contradice ese testimonio. Sin embargo, ¿qué pasaría si se cuestionase a Carla porque, por ejemplo, hubiera otras creencias que sirvan de prejuicio para descartar de antemano su testimonio como lo pudiese ser el machismo y su creencia arraigada de que las mujeres son menos capaces como conocedoras de adquirir y divulgar ciencia? De acuerdo con Miranda Fricker, caeríamos en un caso de injusticia epistémica, en específico, un caso de injusticia testimonial. La injusticia testimonial es definida como un déficit de la credibilidad del hablante debido a un prejuicio identitario (2007, p. 58). Ese prejuicio, contrario a lo que se pueda suponer en principio, no conduce a una discriminación deliberada, sino que más bien es involuntario (2022, p. 20) y puede venir dado por cuestiones meramente transaccionales o por cuestiones estructurales de la interacción (Anderson, 2012).

Esta clase de discriminaciones es especialmente peligrosa por dos motivos: en primer lugar, pone en jaque la posibilidad misma del testimonio, en tanto que se considera a la persona como incapaz de atestiguar veraz, adecuada o atingentemente; en segundo lugar, porque esa misma consideración es un daño a la persona discriminada al tenerle por poco fiable, cuestión que atenta contra su reputación o las consideraciones que esta persona tiene sobre su propia capacidad de conocer.

En cuanto al primer punto, debemos decir lo siguiente: como mencionamos al inicio, el testimonio es la forma más común por la cual las personas adquieren conocimiento. El suponer a alguien como incapaz de participar de esa actividad debido a un prejuicio identitario limita la posibilidad siquiera de considerar al testimonio como una fuente de conocimiento legítimo, pues tendríamos que hablar de cómo ciertos testimonios (aquellos que cumplan con el estándar normativo del que sostiene el prejuicio, sea este el que sea) son dignos de ser tenidos como posible evidencia de que *p*. Esto, como se podrá entrever, suscitaría defender una postura según la cual no caracteriza ese acto de habla como tal el testimonio en cuanto a sus propiedades epistemológicas, sino que simplemente se estaría caracterizando (o más bien, caricaturizando) al agente epistémico al margen del acto de habla que emita, lo que conduciría a otro tipo de análisis, a saber, un análisis erróneo de la supuesta rectitud de lo dicho por las personas al margen de acto de habla emitido.

En cuando al segundo punto, y aún más importante, podemos decir que toda esa consideración epistémica errónea del agente debido a un prejuicio identitario puede derivar en la segregación social debido a la difusión de la reputación negativa, al menos en cuanto a su capacidad de conocedor o conocedora, de la persona afectada. Ese acto de segregación parte del insulto que suscita rechazar el testimonio ajeno (Hazlett, 2017) al presentarse una distribución injusta en la

*economía de la credibilidad*, cuestión que es deriva en un daño a la reputación del agente epistémico en cuanto a su calidad de conocedor (y, extensivamente, a su calidad de persona). Además de las consecuencias sociales esperables de eso (exclusión de grupos, *bullying*, entre otros), esto podría ser tan grave que el mismo agente epistémico termine cuestionando a sí mismo al tener por retroalimentación el rechazo del hablante, lo que conduciría a consecuencias negativas más allá de lo epistémico (falta de confianza en sí mismo/a, autosegregación, depresión, entre otras consecuencias), de modo que no se trata de algo menor. Al contrario, el discriminar la palabra de otros puede tener consecuencias importantes para la vida de esa persona y su relación con el resto.

### **La credibilidad y nuevas tecnologías: *deepfakes***

Para esta última sección, quisiera profundizar en los problemas que suscitan las nuevas tecnologías en relación con la economía de la credibilidad, en específico, el problema que suscitan los *deepfakes* en la discriminación de la palabra de otros en casos derivados de injusticia testimoniales. Para ello quisiera comentar una noticia que impactó a la sociedad chilena y que levanta preocupaciones respecto de las aplicaciones de la IA, pero antes de eso debemos especificar qué entendemos por *deepfake* y en qué medida eso es un problema.

El término *deepfake* se emplea comúnmente para identificar el resultado de emplear inteligencia artificial para la creación de imágenes, videos o audios ultra realistas de personas reales o ficticias. En cuanto el ámbito ético y epistemológico ha habido diversas preocupaciones al respecto, como su rol en la destrucción de la confianza socioepistémica a partir del cuestionamiento de la evidencia en imagen, vídeo o audio (Fallis, 2020; Matthew & Kidd, 2023; Rini, 2020), lo que podría conducir a una suerte de visión apocalíptica de la IA a partir de la idea de que los *deepfakes* son malos per se. Sin embargo, otros señalan que la pregunta acerca de si es una buena aplicación o no la IA es irrelevante, enfatizando la importancia de la regulación de los *deepfakes* en vez de su demonización (Habgood-Coote, 2023). Personalmente estoy a favor lo segundo, pero ello no puede desentenderse de las preocupaciones que suscita lo primero, especialmente si los *deepfakes* pueden ser vehículos de los prejuicios que dan lugar a la discriminación de los testimonios. Como veremos a continuación, el asunto es delicado y relevante tanto en lo epistémico como en lo ético.

A finales de mayo del 2024 se dio a conocer en Chile el caso de estudiantes de educación secundaria que emplearon esta tecnología para generar material pornográfico de sus compañeras de curso. El asunto dio lugar a una batalla legal que terminó con la cancelación de la matrícula de los estudiantes que crearon esos *deepfakes*, pero se denunció que sus compañeras se vieron forzadas a convivir con ellos

durante un proceso que duró meses, cuestión que afectó su integridad (Delgado, 2024). Si bien esta noticia no da cuenta precisamente de un caso de injusticia epistémica y, por tanto, de la discriminación del testimonio de las víctimas, uno puede especular acerca de los posibles desafíos de esta tecnología para ello. Para ello, supongamos el siguiente caso:

### Caso 3: María y los *deepfakes*

Al llegar a clases, María se encuentra con que han difundido *deepfakes* pornográficos con su rostro. Inmediatamente, ella niega ser la persona que aparece en la imagen, pero se topa con los siguientes: 1. Alguno le creen a pesar de la pieza de evidencia que podría ser la foto; 2. Algunos dudan de si creerle debido a la pieza de evidencia que podría ser la foto; 3. Algunos no le creen debido a la pieza de evidencia que podría la foto. Ante la incredulidad de algunos de sus compañeros (los que dudan y los que no le creen), María se siente pasada a llevar, pues considera que no es correcto que no se le crea siendo que ella ha dicho la verdad.

El mejor escenario par María sería (1), pues en él no habría lugar para el daño a su persona. No obstante, tampoco sería culposos (2) y (3) si efectivamente fuese una evidencia la imagen como tal y, en ese sentido, solo hablaríamos de discriminación de información y no de la persona. Sin embargo, he aquí un punto importante: que pueda ser tomada



como evidencia un *deepfake* no lo hace evidencia realmente, de modo que no es equiparable a otra forma de evidencia posible para este contexto como lo podría ser una foto de María. Siendo ese el caso, la consideración errónea del *deepfake* como evidencia muestra un error epistémico de la audiencia al mal evaluar el valor evidencial del testimonio contrario a la supuesta evidencia que sería la imagen. Dado que el testimonio es evidencia potencial del mismo hecho, serviría para derrotar la creencia de que María es quien está realmente en la imagen, mostrando precisamente la negligencia de la audiencia al respecto.

Lo dicho, no parece vincular la injusticia epistémica con el caso necesariamente, sin embargo, podremos ver mejor el punto si presentamos una versión modificada de (3). Supongamos, entonces que:

3.1. Algunos no le creen a María debido un prejuicio machista que les hace considerar al *deepfake* como una pieza de evidencia altamente probable de que ella haya enviado desnudos, lo cual descarta de antemano su testimonio.

Si consideramos (3.1), el cual sería una postura común frente a casos de este estilo, nos encontraríamos con un caso de injusticia testimonial. No quisiera replicar en este artículo la clase de comentarios que uno puede encontrar en RRSS cuando lee noticias como la ya citada, pero el lector podrá

imaginar cuál es el tipo de prejuicio que da lugar a la incredulidad y, por tanto, al daño personal que sufriría María o, en la vida real, las niñas que fueron víctimas de esta difusión de *deepfakes*.

Siendo ese el caso, no podemos simplemente ignorar las posibilidades que tiene el desarrollo de estas nuevas tecnologías en torno a la discriminación arbitraria de la palabra de otros, pues suscita ello daño a su capacidad de como conocedor/a y a su honra como persona. Como ya hemos mencionado, las consecuencias pueden llegar a ser desastrosas para la vida de los involucrados y, lamentablemente, el ámbito de aplicación de los *deepfakes* para afectar la credibilidad y reputación de las personas no se limita solo al ejemplo descrito, sino a cualquier clase de prejuicio identitario que podría servir para ello. Otro ejemplo de esto podría ser la utilización de IA para crear un *deepfake* de una persona afrodescendiente para inculparle de un crimen grave en países donde la discriminación racial es un problema. Al defenderse, esa persona se vería enfrentada a un escenario similar al presentado en (3.1), aunque con consecuencias que podrían afectar tanto su integridad física y psicológica llegando, en el peor de los casos, inclusive al fallecimiento de esa persona (tal y como se podría esperar considerando varios casos de brutalidad policiaca en Estados Unidos como, e.g., lo sucedido con George Floyd en 2020). Como se podrá imaginar, el mismo punto se puede reiterar modificando el prejuicio identitario operando los *deepfakes*,

en este aspecto, como dispositivos que perpetuarían esos prejuicios y, por lo tanto, fomentarían la discriminación de los testimonios basado en el error epistémico que se deriva de la injusticia testimonial.

## Conclusiones

Como hemos visto a lo largo de este artículo, parece haber un vínculo indisoluble entre la epistemología y la ética al estudiar la complejidad que suscita el testimonio, sobre todo considerando el auge de nuevas tecnologías como lo puede ser los *deepfakes*. Se ha llegado a la conclusión de que el contenido multimedia creado con IA puede fomentar las consecuencias negativas de la incredulidad derivada de la injusticia testimonial, mermando aun más la posibilidad de conocer a través de otros por medio del testimonio. Esto, por supuesto, va ligado también a un daño hacia las personas discriminadas, lo cual es una consecuencia ética indeseable de ese tipo de injusticia epistémica por lo que, en definitiva, no se debe desestimar la importancia de examinar la discriminación de los testimonios.

## Referencias

- Adler, J. (2002). *Belief's own ethics*. The MIT Press.
- Anderson, E. (2012). Epistemic justice as a virtue of social institutions. *Social Epistemology*, 26(2), 163–173.
- Anscombe, G. E. M. (2008). What is it to believe someone. En M. Geach & L. Gormally (Eds.), *Faith in a hard*

*ground: Essays on religion, philosophy and ethics* by G.E.M. Anscombe. Imprint Academic. (Obra original publicada en 1979)

Baker, J., & Clark, P. (2018). Epistemic buck-passing and the interpersonal view of testimony. *Canadian Journal of Philosophy*, 48(2), 178–199.

Delgado, F. (2024, 19 de junio). Inteligencia artificial en el Saint George: Acusan que víctimas asisten a clases junto a responsables. *Bio Bio Chile*.  
<https://www.biobiochile.cl/noticias/nacional/region-metropolitana/2024/06/19/inteligencia-artificial-en-el-saint-george-acusan-que-victimas-asisten-a-clases-junto-a-responsables.shtml>

Espinosa, R. (2021). Injusticia epistémica: Consideraciones en torno a la politización de la epistemología del testimonio. En J. A. Valdivia Fuenzalida & R. Espinosa (Eds.), *Ética, creencia y conciencia*. Ediciones Trea.

Fallis, D. (2020). The epistemic threat of deepfakes. *Philosophy and Technology*, 34(4), 623–643.

Faulkner, P. (2011). *Knowledge on trust*. Oxford University Press.

Fricker, E. (2006). Testimony and epistemic autonomy. En J. Lackey & E. Sosa (Eds.), *The epistemology of testimony*. Oxford University Press.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowledge*. Oxford University Press.

- Fricker, M. (2022). Conceptos de injusticia epistémica en evolución. En L. de Brasi & C. Santibáñez Yáñez (Coords.), *Injusticias epistémicas: Análisis y contextos*. Palestra.
- Goldberg, S., & Henderson, D. (2006). Monitoring and anti-reductionism in the epistemology of testimony. *Philosophy and Phenomenological Research*, 72(3), 600–617.
- Habgoode-Coote, J. (2023). Deepfakes and the epistemic apocalypse. *Synthese*, 201(103), 1–23.
- Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7), 335–349.
- Hazlett, A. (2017). On the special insult of refusing testimony. *Philosophical Explorations*, 20(1), 37–51.
- Hinchman, E. (2005). Telling as inviting to trust. *Philosophy and Phenomenological Research*, 70, 562–587.
- Hume, D. (2007). *An enquiry concerning human understanding*. Cambridge University Press.
- Matthews, T., & Kidd, I. J. (2023). The ethics and epistemology of deepfakes. En Fox & Saunders (Eds.), *The Routledge handbook of philosophy and media ethics*. Routledge.
- Moran, R. (2018). Getting told and being believed. En J. Lackey & E. Sosa (Eds.), *The epistemology of testimony*. Oxford University Press.
- Reid, T. (1983). An inquiry into the human mind on the principles of common sense. En R. Beanblossom & K.

- Lehrer (Eds.), *Inquiry and essays*. Hackett. (Obra original publicada en 1764)
- Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosophers' Imprint*, 20(24).
- Ross, A. (1986). Why do we believe what we are told? *Ratio*, 28(1), 69–88.

# ***Deepfake* y regímenes de visibilidad: Lula, Bolsonaro y el gesto inexistente**

*Deepfake and Regimes of Visibility: Lula, Bolsonaro, and the Inexistent Gesture*

**Adriano Charles da Silva Cruz<sup>86</sup>**

## **Resumen**

En abril de 2025, se viralizó un video en el que Lula visitaba a Bolsonaro en un hospital. Creado mediante inteligencia artificial, el video mostraba una escena de reconciliación entre ambos líderes políticos. A través del análisis audiovisual y del estudio de los comentarios publicados en YouTube, este capítulo examina cómo la tecnología deepfake funciona como un dispositivo simbólico

---

<sup>86</sup> Universidade Federal do Rio Grande do Norte.  
<https://orcid.org/0000-0003-2085-302X>

que moviliza afectos, reconfigura los regímenes de visibilidad y contribuye a las disputas ideológicas contemporáneas. A pesar de su carácter ficticio, la escena activó interpretaciones contradictorias, marcadas por el humor, la moral cristiana, la repulsión o el deseo de pacificación. Se concluye que las deepfakes no solo amenazan la confianza pública, sino que también articulan deseos y tensiones del imaginario social, desestabilizando la frontera entre lo posible y lo real.

**Palabras clave:** Inteligencia artificial; *deepfake*; visibilidad; posverdad; comunicación política.

## **Abstract**

In April 2025, a video went viral showing Lula visiting Bolsonaro in a hospital. Created using artificial intelligence, the video depicted a scene of reconciliation between the two political leaders. Through audiovisual analysis and a study of comments posted on YouTube, this chapter explores how deepfake technology functions as a symbolic device that mobilizes affects, reshapes regimes of visibility, and contributes to contemporary ideological disputes. Despite its fictional nature, the scene elicited contradictory interpretations, marked by humor, Christian morality, repulsion, or the desire for pacification. It is concluded that deepfakes not only threaten public trust but also articulate desires and tensions within the social imaginary, destabilizing the boundary between the possible and the real.



**Keywords:** Artificial intelligence; deepfake; visibility; post-truth; political communication.

## Introducción

A principios de abril de 2025, comenzó a circular en redes sociales un breve video, de aproximadamente 48 segundos, en el que el presidente Lula aparece visitando a Jair Bolsonaro en un hospital. El mandatario se acerca, toca la frente de su adversario, le toma la mano y finalmente lo abraza con fuerza. La imagen es hiperrealista: el entorno hospitalario, la iluminación, el vestuario del paciente y del visitante, todo parece auténtico. Pero el abrazo nunca ocurrió. Se trata de una *deepfake*, publicada en el canal Belzontemovies en YouTube y rápidamente compartida en grupos de WhatsApp. En el breve lapso entre su publicación y la redacción de este artículo, el video ya había superado las 400 mil visualizaciones y acumulado 403 comentarios.

Este acercamiento entre los dos líderes no es un caso aislado: en diciembre de 2024, la agencia de verificación de datos Aos Fatos<sup>87</sup> desmintió otra escena hospitalaria, en este caso, a la inversa—, en la que Lula, supuestamente internado por un accidente doméstico, habría recibido la visita del expresidente (Bortolon, 2024). En abril de 2025 también

---

<sup>87</sup> Aos Fatos es una agencia brasileña independiente de verificación de hechos, especializada en combatir la desinformación y evaluar la veracidad de discursos públicos.

circuló una imagen en la que Lula habría conmovido a Bolsonaro al desearle una "pronta recuperación", según reportó UOL Confere<sup>88</sup> (Espina, 2025).

La proliferación de imágenes fabricadas con alto grado de realismo, como las *deepfakes*, debe entenderse a partir de las transformaciones del ecosistema mediático: la disolución de la frontera entre emisores y audiencias. A esto se suma la lógica de circulación acelerada de imágenes en redes sociales digitales, mediada por algoritmos que tienden a privilegiar el conflicto y la polémica (Levy et al., 2023).

En un contexto de conflicto persistente, las imágenes de reconciliación fabulan un momento de distensión simbólica. En un país que avanza hacia una mayoría evangélica, especialmente en su vertiente neopentecostal<sup>89</sup>, la simulación de un perdón o un encuentro fraterno podría constituir un atractivo emocional para parte del público. Al mismo tiempo, estas montajes pueden profundizar la desconfianza hacia la política representativa.

---

<sup>88</sup> UOL Confere es una plataforma de verificación de datos vinculada al principal portal de noticias de Brasil, UOL (Universo Online).

<sup>89</sup> En Brasil, el crecimiento del cristianismo evangélico —especialmente las iglesias neopentecostales— ha tenido fuerte impacto en la política, promoviendo una agenda conservadora y articulando una base electoral significativa.

En esa dirección se orienta nuestra hipótesis: las deepfakes, más allá de su efecto ilusorio, contribuyen al recrudecimiento del antagonismo social entre el bolsonarismo y el lulopetismo<sup>90</sup> al promover un gesto de apaciguamiento que nunca ocurrió, y que, para muchos, nunca debería ocurrir.

Las tecnologías de inteligencia artificial reconfiguran los regímenes de visibilidad y de creencias. Con las *deepfakes*, rostros familiares comienzan a decir lo que nunca dijeron y, sin embargo, siguen siendo reconocibles. Esa familiaridad artificial les confiere potencia: no necesitan ser perfectas para generar efecto. Al activar un recuerdo, una creencia preexistente o, sobre todo, un afecto ya presente, desencadenan procesos de identificación o de rechazo.

Este tipo de producción audiovisual aporta beneficios en áreas como el cine, la publicidad o la educación, donde la simulación sirve a la creatividad, la reconstrucción o la experimentación (Weerawardana & Fernando, 2021). No obstante, también produce efectos negativos, como fraudes financieros, difusión de pornografía no consensuada y

---

<sup>90</sup> Bolsonarismo y lulopetismo son términos populares en Brasil que describen, respectivamente, a los seguidores de Jair Bolsonaro y a los de Luiz Inácio Lula da Silva (y el Partido de los Trabajadores - PT), reflejando una polarización política extrema.

estrategias de desinformación, especialmente en contextos electorales (Mirsky & Lee, 2021; Tolosana et al., 2020).

Durante la campaña para la alcaldía de Buenos Aires en 2025, circuló en redes un video *deepfake* que simulaba al expresidente Mauricio Macri declarando su apoyo al candidato Manuel Adorni. La pieza fue utilizada para confundir a los votantes y debilitar la candidatura de Silvia Lospennato, del PRO. Aunque el video fue retirado por orden de la Corte Electoral, el daño ya estaba hecho. Manuel Adorni (La Libertad Avanza) fue electo con el 30% de los votos. El episodio reavivó el debate sobre la regulación de las redes sociales y el uso de tecnologías de manipulación audiovisual en la competencia política y en el propio proceso democrático. Este caso ilustra cómo las tecnologías de generación sintética pueden intervenir directamente en disputas electorales contemporáneas, reforzando los escenarios de desinformación descritos en la literatura sobre posverdad.

En Bolivia, durante el proceso electoral de 2025, las redes sociales también se convirtieron en un espacio central para la circulación de contenidos manipulados. Videos alterados, audios editados y publicaciones fuera de contexto circularon ampliamente en plataformas digitales, afectando el debate público. Entre los casos reportados se encuentra la difusión de videos que atribuían declaraciones falsas a candidatos y líderes políticos, así como audios manipulados que distorsionaban posiciones públicas. De acuerdo con un

reportaje de El País, la campaña electoral boliviana estuvo marcada por la viralización de contenidos desinformativos amplificados por páginas digitales y creadores de contenido con alta audiencia, lo que contribuyó a intensificar la circulación de narrativas engañosas durante el período electoral (Ruvenal, 2025). Estos episodios muestran cómo la lógica de viralización de las plataformas digitales puede potenciar la circulación de imágenes y discursos manipulados, afectando la percepción pública de los actores políticos y la confianza en el proceso democrático.

En Brasil, las *deepfakes* ganan fuerza precisamente donde la disputa es más intensa: en el campo político. Lula ha sido blanco de videos falsos vinculados a estafas bancarias (Secretaria de Comunicação Social da Presidência da República, 2023), tuvo su imagen manipulada en materiales de campaña que fueron eliminados por decisión judicial (Bergamo, 2024), e incluso apareció en escenas ficticias, como un video navideño en el que compartía una celebración con Bolsonaro y otras figuras públicas, revelando la ambigüedad de este tipo de imágenes (Moretti, 2025).

Ante la proliferación de estas imágenes y sus consecuencias, el gobierno creó un grupo interministerial para debatir el uso ético de la inteligencia artificial (Poder360, 2025), mientras el Congreso discute propuestas para considerar el uso de imágenes sintéticas como forma de violencia (Castro, 2025). El Tribunal Electoral, por su parte,

enfrenta dificultades para rastrear y frenar el uso de estas imágenes en campañas destinadas a deslegitimar adversarios (Gama & Nani, 2025).

Este capítulo tiene por objetivo analizar el video “Lula deseando mejoría a Bolsonaro internado en el hospital”, su *mise-en-scène* y diálogos, disponible en YouTube. Se busca comprender cómo las *deepfakes* generan debates, movilizan afectos e instalan disputas en las redes sociales.

Tras una revisión bibliográfica y documental, se procedió al análisis del objeto empírico, seleccionado por su volumen de interacciones en un corto periodo, así como por las diversas reacciones que provocó en la audiencia. El análisis se organizó a partir de la descripción e interpretación del producto audiovisual.

Para comprender mejor la circulación de esta imagen, se realizó un estudio exploratorio de los primeros cien comentarios publicados en la plataforma. Esta elección responde a la lógica del ordenamiento algorítmico, que tiende a priorizar las publicaciones con mayor interacción, comentarios, número de “me gusta”, respuestas y tiempo de visualización.

Es importante señalar que los comentarios fueron transcritos sin correcciones ortográficas ni gramaticales, con el fin de preservar su forma original. Por razones éticas, los

nombres de los autores fueron suprimidos, garantizando el anonimato de los participantes, conforme a las directrices para investigaciones con datos públicos en plataformas digitales. Estas interacciones aportan elementos valiosos para comprender las apropiaciones afectivas y los atravesamientos ideológicos provocados por la pieza.

## **Fundamentación teórica**

Las *deepfakes* adquieren fuerza en la era de la “posverdad”, caracterizada por la preeminencia de la desinformación y por el debilitamiento de la racionalidad (D’Ancona, 2018). En este contexto, lo que importa ya no es necesariamente lo verdadero, sino aquello que resulta plausible o que confirma lo que el público ya cree. Este fenómeno privilegia las emociones y las creencias, transformando las redes sociales digitales en plataformas clave para la difusión de contenidos manipulados. Según Bakir y McStay (2024), los afectos no solo influyen en el debate público, sino que también modelan la propia percepción de lo que se entiende por realidad.

Desde esta perspectiva, Sodr  (2006) sostiene que los medios de comunicaci n no se limitan a transmitir mensajes, sino que act an como dispositivos que movilizan afectos, deseos y sensaciones en la constituci n de v nculos sociales. En otro momento, el autor (Sodr , 2019) afirma que vivimos bajo el dominio de un “continuo medi tico”, en el cual el

discurso político ya no se apoya principalmente en la racionalidad argumentativa, sino en operaciones simbólicas marcadas por la repetición, el impacto sensorial y la saturación.

A ello se suma el papel de la gubernamentalidad algorítmica, que favorece la circulación de contenidos conflictivos. En un estudio empírico, Levy et al. (2023) demostraron que los algoritmos de las redes sociales amplifican publicaciones cargadas de afectos negativos y hostilidad. Aunque dichos contenidos no coincidan con las preferencias declaradas de los usuarios, son priorizados por generar mayor compromiso.

La mentira, en este caso, no aparece como una ruptura puntual, sino como un flujo constante, diluido entre memes, montajes, rumores y gestos espectaculares. La crítica se debilita, pues la duda deja de ser una etapa del pensamiento y se convierte en un estado permanente de suspensión.

Además, es necesario considerar las estructuras sociales que organizan la circulación de las creencias. Nguyen (2020) distingue entre “epistemic bubbles” y “echo chambers”, dos configuraciones informativas que suelen confundirse en el debate público. Mientras que las primeras se caracterizan por la simple omisión de voces disonantes en el flujo informativo, las segundas implican un proceso más radical: la deslegitimación activa de las fuentes externas. En



una burbuja epistémica, la exposición a información contradictoria puede corregir creencias erróneas; en cambio, en una cámara de eco, toda información proveniente del exterior es previamente desacreditada. Bajo estas condiciones, incluso las evidencias o advertencias técnicas sobre la manipulación de una imagen pueden ser ignoradas o reinterpretadas como parte de una supuesta conspiración. Así, el problema no se limita a la falsedad de los contenidos, sino a las redes de confianza que estructuran su interpretación social.

Desde esta perspectiva, las “echo chambers” descritas por Nguyen (2020) no solo limitan la exposición a información divergente, sino que también reorganizan los regímenes de visibilidad (Thompson, 1995), definiendo qué imágenes y enunciados pueden ser reconocidos como legítimos o sospechosos dentro de comunidades discursivas polarizadas.

No obstante, es importante reconocer que no todo contenido manipulado tiene la intención deliberada de engañar. Wardle y Derakhshan (2017) proponen una tipología precisa de las formas de desorden informacional: misinformation (error sin intención), disinformation (manipulación intencionada) y malinformation (uso dañino de contenidos verídicos).

Es en este escenario de creciente complejidad que emergen las *deepfakes*, capaces de producir escenas con alto grado de realismo, simulando con tal fidelidad que pueden engañar incluso a las miradas más entrenadas. Una tecnología que antes estaba restringida a especialistas, hoy se vuelve cada vez más accesible gracias al avance de la inteligencia artificial generativa, que permite crear imágenes y videos mediante simples comandos de texto.

Según Masood et al. (2021), existe una proliferación de herramientas disponibles para la generación de videos y audios sintéticos, muchas de ellas con interfaces simples, gratuitas y de amplia difusión. Ya no es necesario ser un experto para producir contenidos que simulen con precisión gestos, voces y expresiones. Mirsky y Lee (2020) advierten que, a medida que estas tecnologías avanzan, los filtros de detección se tornan rápidamente obsoletos.

Por otro lado, es necesario considerar los usos constructivos de esta tecnología. Misirlis y Munawar (2023) discuten la posibilidad de emplear *deepfakes* de manera ética en contextos educativos o artísticos, como en el cine, siempre que haya claridad, consentimiento y una advertencia explícita de ficcionalización.

En Brasil, no faltan ejemplos de la apropiación de *deepfakes* en contextos de tensión política. Durante los ciclos electorales más recientes, circularon intensamente videos

manipulados, muchas veces vinculados a campañas de desinformación. Souza (2023) analiza los dilemas legales y morales que plantea este tipo de material, alertando sobre la velocidad con la que se disemina y el impacto que puede generar en la confianza institucional.

Más recientemente, Magno y Pieroni (2021) señalaron que las deepfakes ya están siendo utilizadas estratégicamente para la fabricación de noticias falsas con fines electorales. En estos casos, la disputa trasciende la esfera individual y amenaza directamente la integridad del proceso democrático. Cuando circulan pruebas falsas como si fueran auténticas, la propia noción de evidencia entra en colapso.

En síntesis, las *deepfakes* no se limitan a la invención de mentiras visuales, sino que operan como instrumentos de construcción de versiones del mundo que resuenan con deseos latentes en el tejido social. El truco no reside únicamente en la técnica, sino en el afecto que convoca: afectos capaces de movilizar resentimientos, reforzar creencias preexistentes y estimular conflictos.

De este modo, es fundamental comprender los regímenes de visibilidad (Thompson, 1995), es decir, los modos en que una imagen se vuelve perceptible y creíble, y bajo qué condiciones es reconocida como legítima. Toda imagen se inscribe en una dinámica social, produce significados y moviliza afectos.

Verón (1995), por su parte, propone la noción de dispositivo de enunciación para dar cuenta de la maquinaria simbólica que organiza la emergencia de estas imágenes: los códigos lingüísticos, los formatos técnicos y los pactos de lectura. En el caso de las deepfakes políticas, estos dispositivos son cuidadosamente contruidos para simular familiaridad, y así suspender la duda.

### **Imagen de una reconciliación fabulada**

La escena tiene apenas 48 segundos y transcurre íntegramente en una habitación de hospital. Jair Bolsonaro aparece acostado, con expresión serena y vestido con ropa de internación. Lula, con traje oscuro y corbata, se acerca lentamente al lecho. Con semblante tranquilo, le toca suavemente la frente a Bolsonaro y, en seguida, le toma la mano con delicadeza. El momento más intenso ocurre cuando Lula lo abraza con fuerza, apoyando su rostro en el hombro del otro, en un gesto afectuoso. Bolsonaro, con los ojos entrecerrados, lo recibe con una leve sonrisa. El entorno es minimalista: paredes blancas, grandes ventanas con luz natural y algunos equipos electrónicos al fondo<sup>91</sup>.

---

<sup>91</sup> Video disponible en *YouTube*: “Lula deseando mejoría a Bolsonaro internado en el hospital”, publicado en el canal *Belzontemovies*. Disponible en: <https://www.youtube.com/watch?v=pZQn8pSyupM> (consultado el 5 de enero de 2026).

La fotografía está compuesta por una secuencia estática y sin cortes. Se utiliza un plano medio abierto, frontal, que encuadra la cama con ambos personajes en escena: Lula de pie a la derecha, Bolsonaro acostado a la izquierda, con el rostro visible. Lula ocupa la mayor parte del campo visual.

A pesar del realismo visual, la mise-en-scène conserva una estructura narrativa y una gestualidad algo artificial, lo que delata la existencia de un guión ensayado. El diálogo simulado entre los dos personajes incluye frases como:

— Uh... hola... opa... entra.

— Ay, mi amigo, ven aquí para que te abrace.

— ¿Qué es esto, compañero? ¿Qué te pasó? ¿Estás bien?

— Estoy bien, sí. Solo un pequeño dolor en la barriga... pero pronto estaré mejor y volveré a casa.

— ¡Qué susto, Bolsonaro!

La puesta en escena audiovisual funciona como un dispositivo de enunciación que construye una reconciliación imaginada, elaborada mediante desplazamientos ideológicos significativos. Más adelante, Lula, con tono humorístico, comenta:

— Cancelé todos mis compromisos. Le di mi tarjeta corporativa a Janjinha<sup>92</sup> para que la gastara y así poder venir a verte.

La inclusión de este enunciado activa una red de referencias recientes, como las denuncias sobre el uso de tarjetas del gobierno por parte de Lula, y el apodo utilizado públicamente para referirse a la primera dama. Este entrelazamiento entre lo íntimo y lo político intensifica el carácter performativo de la deepfake, atravesada por el humor y la memoria.

La escena prosigue con un intercambio afectivo: Lula ofrece una botella de whisky y la conversación se desarrolla en un tono amistoso. El diálogo concluye con un nuevo abrazo:

— Qué bueno que estás mejor. Que solo fue un susto, y que pronto volveremos a nuestra vida normal, mi amigo.

— Tú sabes que eres muy importante para mí.

— Muchas gracias.

---

<sup>92</sup> En 2023, surgieron críticas por supuestos abusos en el uso de la “tarjeta corporativa” por parte de la primera dama —un instrumento destinado a cubrir gastos oficiales del gobierno brasileño—, lo que generó repercusión mediática y política.

— No te preocupes. Puedes estar seguro de que no te vas a librar de mí tan fácilmente.

— ¿Lo prometes?

— Dame otro abrazo.

— Descansa bien... y cualquier cosa, mándame un “zap”<sup>93</sup> que vengo a verte.

Este gesto de apaciguamiento evoca una relación de dependencia simbólica mutua entre las figuras políticas. Expresiones como “compañero” y “mi amigo” refuerzan el tono de intimidad. El video ironiza sobre esta co-dependencia: los antagonismos se alimentan entre sí. Como se sabe, la elección de Lula en 2022 estuvo marcada por el rechazo a Bolsonaro. A su vez, el bolsonarismo ganó fuerza como reacción simbólica al lulismo.

Lula obsequia a Bolsonaro una bebida alcohólica. El gesto rompe con la expectativa tradicional de una visita hospitalaria, promoviendo la risa. También actualiza la imagen de un presidente informal, bohemio, transgresor de las normas institucionales.

---

<sup>93</sup> Zap” es una forma coloquial de referirse a *WhatsApp* en Brasil.

Finalmente, en la despedida, emerge el campo religioso como marcador de afectividad:

— Quédate con Dios.

— Igualmente. Que así sea.

Este reordenamiento fraterno solo es posible porque el video funciona como un dispositivo técnico y simbólico de desplazamiento ideológico: las imágenes combativas de ambos líderes rivales son reorganizadas bajo otro régimen de visibilidad, en el que ahora son interpretados como amigos.

Los avatares son contruidos con voces pausadas, gestos suaves y un vocabulario afectivo. Se trata, por tanto, de una escenificación mediática que invita al público a percibir no solo a líderes políticos, sino a figuras familiares.

En el caso específico de Lula, esta ya es una imagen conocida: el “Lula paz y amor” de la campaña de 2002<sup>94</sup> reaparece aquí en otro contexto histórico.

Esta operación mediática se articula con una estrategia de circulación en redes sociales mediante hashtags que amplifican su alcance y capacidad de engagement. El video

---

<sup>94</sup> El lema “Lula paz e amor” fue una estrategia de marketing político utilizada en la campaña presidencial de 2002 para suavizar la imagen radical de Lula y atraer votantes moderados.



fue lanzado con 48 términos, incluyendo 18 repeticiones en variantes de mayúsculas y minúsculas. Algunas de las palabras clave (#Bolsonaro, #Lula, #luisinacio, #jairbolsonaro, #lula2026, #bolsonaro2026, #bolsolula, #inelegível, #anistiajá, #presidente, #política) refuerzan la polarización entre ambos adversarios, generando un ciclo de visibilidad.

Otros hashtags refuerzan los sentidos conciliadores contruidos por el video (#tiempo, #paz, #visitaInesperada, #salud, #reunión, #visita, #urgencia, #cama, #barriga). También se incluyen etiquetas que indican la autoría y la técnica utilizada (#AI, #Belzontemovie, #belzontemovie11). Finalmente, aparecen enunciados orientados a captar atención en YouTube y movilizar a la audiencia (#urgente, #corre, #sucedió, #viral, #viralizado).

La inscripción “tem[p]o de paz”, reiterada al final del subtítulo del video, sintetiza el tono conciliador propuesto, pero también activa capas de ironía, dada la improbabilidad de un encuentro real en los términos dramatizados. Aunque el aviso de contenido sintético está presente en una pequeña pantalla inicial, el realismo técnico de la imagen y el tono de la locución refuerzan la invitación a la creencia.

En este punto, la *deepfake* funciona como una pieza performativa que inscribe un gesto inexistente, pero

imaginable. Actúa como un artefacto que reconfigura los modos de visibilidad de los sujetos políticos contemporáneos.

La imagen de los políticos, anteriormente asociada a registros documentales o periodísticos, pasa a habitar un territorio donde lo real y lo simulado se entrecruzan. La técnica importa, pero las pasiones movilizadas posibilitan interpretaciones divergentes.

Aquí, la *deepfake* opera menos como engaño y más como deseo, repulsión o imposibilidad.

### **La audiencia y las disputas de interpretación**

En los comentarios, la audiencia interpreta, responde, se ve afectada y produce sentidos diversos. Algunos leen el video como una alegoría de un país deseado, proyectando en él valores cristianos, lo que reafirma el carácter simbólico del montaje como catalizador de afectos. Otros tensionan la escena mediante el sarcasmo, la incredulidad o la repulsión, revelando los límites de la verosimilitud y de la aceptación de la fabulación política.

En ciertos comentarios se identifica un deseo de pacificación nacional. Son enunciadores que utilizan la puesta en escena para proyectar una versión idealizada de Brasil, en la cual los polos políticos opuestos superan sus diferencias en nombre del bien común. Frases como “Sería bueno si todos fueran así, todo el mundo en paz”, “Lo correcto sería que el

amor entre las personas fuera incondicional...” o “Eso es justamente lo que deberíamos hacer, así habría la tan soñada paz mundial 🌿🙏” indican una lectura de la imagen como alegoría de un país reconciliado, anclado en una moral cristiana.

Este marco simbólico se refuerza aún más en enunciados que interpretan la escena desde una ética evangélica del amor al prójimo. Comentarios como “Me gustó, la Biblia dice que el primer mandamiento es el amor. Jesús manda amar unos a otros”, “La Biblia habló, pero Lucifer no combina con la honestidad” y “Este es el verdadero mensaje de Jesús. Ojalá fuera verdad” ejemplifican la transversalidad del campo religioso en lo político-partidario. Tales comentarios construyen la imagen de “creyentes” que buscan, en la política, signos de reconciliación basados en enseñanzas evangélicas.

Por otro lado, aparece un conjunto de comentarios marcados por el escepticismo y la ironía, que señalan la naturaleza artificial del video y expresan incredulidad ante la escena. Frases como “¿Esto es un montaje!? Qué bueno sería si fuera cierto”, “Engáñame que me GUSTA 😁” deslegitiman la escena y la insinúan como manipulación visual. Aquí, la técnica del *deepfake* no solo es reconocida, sino también cuestionada como amenaza a la autenticidad del gesto político.

Ese escepticismo, en muchos casos, da paso al resentimiento y a una polarización ideológica explícita. Varios comentarios revelan rechazo tanto hacia el montaje como hacia la posibilidad misma de un acercamiento entre ambos políticos. Expresiones como “¡Más respeto, por favor!”, “Ese borracho es una comedia 😂😂😂” y “¡YO ENFERMO Y SATANÁS VINIENDO A VISITARME! TRÁGICO...” o “Cuando se trata de Lula, ponen a Bolsonaro entrando disfrazado de enfermera agarrando el cuello de Lula; cuando es Bolsonaro, ponen a Lula entrando con todo el amor del mundo. Escenas de amor y odio. Siendo que Lula es amor y Bolsonaro es odio” articulan repulsión, ironía agresiva e indignación moral. Estos enunciados refuerzan la idea de que el antagonismo entre Lula y Bolsonaro es irreconciliable, y que la escena de la visita hospitalaria debe leerse como una provocación por parte del petismo.

En otra dirección, surgen comentarios que movilizan el humor, el sarcasmo o la despolitización como formas de enfrentar el contenido: “Me estoy muriendo de risa”, “Jajaja. Esta es buena. ¡Como si fuera cierto!” y “¡Sabía que era inteligencia artificial! jajajajajaja”.

Asimismo, mensajes como “Ahí están el sucio y el más sucio, esta es nuestra política y todavía puede empeorar”, “Esos dos se están riendo de los payasos que son el pueblo” o “Son iguales, deberían dejar a sus esposas y vivir juntos” refuerzan el desencanto con la política partidaria.

En ese sentido, el video analizado no sólo escenifica una posibilidad improbable, sino que también cataliza una miríada de reacciones afectivas e ideológicas que ilustran los desplazamientos de la comunicación política contemporánea. Mediante la simulación técnica, se construye un escenario de negociación simbólica en el que lo imposible se presenta como verosímil, y el público es convocado a ocupar activamente la escena: sea para desear, creer, dudar, burlarse o rechazar.

## **Consideraciones finales**

El análisis de la narrativa evidenció cómo recursos técnicos propios de la inteligencia artificial, como la simulación de voz, las expresiones faciales y la escenografía, fueron movilizados para construir una escena de reconciliación improbable entre antagonistas políticos. Aunque reconocidamente ficticia, la escena opera bajo una lógica de verosimilitud que activa reacciones afectivas intensas.

El video *deepfake* actúa como un dispositivo de enunciación que, al simular intimidad y emoción mediante códigos mediáticos, desestabiliza el imaginario político. El montaje desplaza a los sujetos políticos y convoca a la audiencia, por medio de hashtags, a reaccionar desde los afectos.

En los comentarios se observa que la audiencia construye sentidos diversos, a veces contradictorios entre sí: reconciliación, escepticismo, sátira y resentimiento. Algunas interpretaciones se apoyan en criterios objetivos, como el reconocimiento técnico del montaje, mientras que otras emergen de la fuerza de los afectos y de convicciones personales, incluso religiosas. En este campo de disputas, los discursos oscilan entre la desconfianza, la ironía y el sarcasmo.

El caso analizado demuestra que las deepfakes, lejos de ser meras herramientas de desinformación o sátira, funcionan como dispositivos mediáticos capaces de movilizar debates públicos y generar interés social. La reconciliación escenificada, presentada con matices de humor, reaparece en los comentarios como un deseo latente por parte de la audiencia, muchas veces atravesado por la moral cristiana que estructura parte del imaginario social brasileño.

Comprender las dinámicas sociales provocadas por estas imágenes resulta fundamental para reflexionar sobre los desafíos de la democracia en tiempos de crisis, cuando la frontera entre lo posible y lo real se convierte cada vez más en una zona de disputa simbólica y afectiva.

Los resultados obtenidos permiten pensar que este tipo de dinámicas no se limita al caso brasileño. En contextos latinoamericanos marcados por una fuerte polarización

política, es posible imaginar situaciones similares en las que imágenes sintéticas simulen gestos improbables entre adversarios históricos o atribuyan declaraciones inexistentes a líderes políticos durante campañas electorales. Un video que mostrara, por ejemplo, a dos candidatos rivales escenificando un acuerdo inesperado o una confesión política inexistente podría circular rápidamente en redes sociales, activando interpretaciones divergentes entre escepticismo, humor, indignación o esperanza. Incluso cuando tales imágenes son posteriormente desmentidas, su circulación inicial puede contribuir a reorganizar temporalmente los marcos de interpretación pública. Este tipo de escenarios permite comprender que las *deepfakes* no solo fabrican falsedades visuales, sino que operan dentro de los regímenes de visibilidad contemporáneos, donde la verosimilitud emocional y la circulación algorítmica pueden producir efectos políticos antes incluso de que la veracidad de la imagen sea cuestionada.

## Referencias

- Bakir, V., & McStay, A. (2024). Fake news in the age of AI: The impact of emotion, trust, and data voids [Publicación anticipada en línea]. *Journal of Information Technology*.
- Bergamo, M. (2024, 5 de septiembre). Justiça Eleitoral manda Marçal excluir vídeo com deepfake de Boulos e Lula. *Folha de S.Paulo*.

<https://www1.folha.uol.com.br/colunas/monicabergamo/2024/09/justica-eleitoral-manda-marcal-excluir-video-com-deepfake-de-boulos-e-lula.shtml>

Bortolon, B. (2024, 12 de diciembre). Imagem de Bolsonaro visitando Lula no hospital foi gerada por IA. *Aos Fatos*.

<https://www.aosfatos.org/noticias/imagem-inteligencia-artificial-bolsonaro-lula-hospital/>

Castro, L. E. de. (2025, 24 de abril). "Deepfake" com uso de IA e discriminação contra gestantes: Lula sanciona leis de proteção às mulheres; veja o que muda. *O Globo*.

<https://oglobo.globo.com/brasil/noticia/2025/04/24/deepfake-com-uso-de-ia-e-discriminacao-contragestantes-lula-sanciona-leis-de-protecao-as-mulheres-veja-o-que-muda/>

D'Ancona, M. (2018). *Pós-verdade: A nova guerra contra os fatos em tempos de fake news* (M. E. Ávila, Trad.). Faro Editorial.

Espina, R. (2025, 23 de abril). É falso que Lula tenha desejado recuperação a Bolsonaro e que ele tenha chorado; vídeo e narrativa são enganosos. *UOL Confere*.

<https://noticias.uol.com.br/confere/ultimas-noticias/2025/04/23/mensagem-lula-recuperacao-bolsonaro-choro-falso.htm>

Gama, R., & Nani, L. (2025, 15 de junio). Manipulação digital: Deepfake ultrarrealista eleva preocupação com eleição e desafia fiscalização para 2026. *O Globo*.

<https://oglobo.globo.com/blogs/sonar-a-escuta-das->



[redes/post/2025/06/manipulacao-digital-deepfake-ultrarrealista-eleva-preocupacao-com-eleicao-e-desafia-fiscalizacao-para-2026.ghtml](https://redes/post/2025/06/manipulacao-digital-deepfake-ultrarrealista-eleva-preocupacao-com-eleicao-e-desafia-fiscalizacao-para-2026.ghtml)

Levy, R., Barberá, P., Gill, P., González-Bailón, S., Hilbert, M., & Watts, D. J. (2023). *Engagement, user satisfaction, and the amplification of divisive content on social media*. Knight First Amendment Institute at Columbia University.

<https://knightcolumbia.org/content/engagement-user-satisfaction-and-the-amplification-of-divisive-content-on-social-media>

Magno, J. C., & Pieroni, G. M. (2021). Os perigos do deepfake para a democracia brasileira. En *V Seminário Internacional de Pesquisa em Midiatização e Processos Sociais: Anais*. Universidade Federal de Santa Maria.

<https://www.midiaticom.org/anais/index.php/seminario-midiatizacao-artigos/article/view/1748>

Masood, M., Yousaf, M. H., Muhammad, K., Kwon, S., & Baik, S. W. (2021). *Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward* (arXiv:2103.00484). arXiv.

<https://arxiv.org/abs/2103.00484>

Mirsky, Y., & Lee, W. (2020). The creation and detection of deepfakes: A survey. *ACM Computing Surveys*, 54(1), Article 7. <https://doi.org/10.1145/3425780>

- Misirlis, N., & Munawar, H. B. (2023). *From deepfake to deep useful: Risks and opportunities through a systematic literature review* (arXiv:2311.15809). arXiv. <https://doi.org/10.48550/arXiv.2311.15809>
- Moretti, F. (2025, 16 de junio). Deepfake ultrarrealista eleva preocupação com eleição e desafia fiscalização para 2026. *Mais Goiás*. <https://www.maisgoias.com.br/brasil/deepfake-ultrarrealista-eleva-preocupacao-com-eleicao-e-desafia-fiscalizacao-para-2026/>
- Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161. <https://doi.org/10.1017/epi.2018.32>
- Poder360. (2025, 12 de mayo). *Governo cria grupo para gerir plano de inteligência artificial*. <https://www.poder360.com.br/poder-governo/governo-cria-grupo-para-gerir-plano-de-inteligencia-artificial/>
- Ruvenal, C. (2025, 11 de agosto). Videos que desinforman: La campaña sucia toma las redes sociales en Bolivia. *El País*. <https://elpais.com/america/2025-08-11/videos-que-desinforman-la-campana-sucia-toma-las-redes-sociales-en-bolivia.html>
- Secretaria de Comunicação Social da Presidência da República. (2023, 11 de diciembre). *Estelionatários usam imagem do presidente da República em deep fake*. Gov.br. <https://www.gov.br/secom/pt->

[br/fatos/brasil-contrafake/noticias/2023/12/estelionatarios-usam-imagem-do-presidente-da-republica-em-deep-fake](https://br/fatos/brasil-contrafake/noticias/2023/12/estelionatarios-usam-imagem-do-presidente-da-republica-em-deep-fake)

- Sodré, M. (2006). *As estratégias sensíveis: Afeto, mídia e política*. Vozes.
- Sodré, M. (2019). O facto falso: Do factóide às fake news. En J. Figueira & S. Santos (Orgs.), *As fake news e a nova ordem (des)informativa na era da pós-verdade* (pp. 87–100). Imprensa da Universidade de Coimbra.
- Souza, V. G. P. de. (2023). Fake news política no Brasil: Da ficha falsa da Dilma à deepfake do Jornal Nacional. *Revista Alterjor*, 28(2), 591–602. <https://doi.org/10.11606/issn.2176-1507.v28i2p591-602>
- Thompson, J. B. (1995). *A mídia e a modernidade: Uma teoria social da mídia* (M. C. B. Miceli, Trad.). Vozes.
- Tolosana, R., Vera-Rodríguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131–148. <https://doi.org/10.1016/j.inffus.2020.07.007>
- Verón, E. (1995). *Fragments de um discurso audiovisual*. Brasiliense.
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe.

<https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>

Weerawardana, A., & Fernando, A. (2021). Deepfake technology: Applications, challenges, and legal issues. En *2021 IEEE International Conference on Computer Communication and the Internet (ICCCI)* (pp. 124–129). IEEE.

<https://doi.org/10.1109/ICCCI54341.2021.00030>

**“I don't like them putting  
fraud in my conspiracy  
theories!”: A  
Philosophical  
Examination of  
Conspiracy Theories,  
‘Pataphysics, and Alex  
Jones’s Frog Conspiracy<sup>95</sup>**

*“¡No me gusta que metan fraude en mis teorías  
conspirativas!”: un examen filosófico de las  
teorías conspirativas, la ‘patafísica y la  
Conspiración de las Ranas de Alex Jones*

---

<sup>95</sup> The original title was adjusted to comply with the terms established by the publisher and the call for papers. The authors wish to note their preference for the initial version.

**Javier Silva Silva<sup>96</sup>, Gabriel Donoso Umaña<sup>97</sup>, Esteban Echaniz<sup>98</sup>**

*This is the fatality of every system committed by its own logic to total perfection and therefore to a total defectiveness, to absolute infallibility and therefore irrevocable breakdown: the aim of all bound energies is their own death. This is why the only strategy is catastrophic, and not dialectical at all. Things must be pushed to the limit, where quite naturally they collapse and are inverted. At the peak of value, we are closest to ambivalence, at the pinnacle of coherence we are closest to the abyss of corruption which haunts the reduplicated signs of the code. Simulation must go further than the system. Death must be played against death: a radical tautology that makes the system's own logic the ultimate weapon. The only strategy against the hyperrealist system is some form of pataphysics, 'a science of imaginary solutions'; that is, a science-fiction of the system's reversal against itself at the extreme limit of simulation, a*

---

<sup>96</sup> Universidad de Santiago de Chile. <https://orcid.org/0000-0002-7474-3224>

<sup>97</sup> Universidad de Santiago de Chile. <https://orcid.org/0000-0002-4842-1387>

<sup>98</sup> Universidad de Santiago de Chile. <https://orcid.org/0000-0002-6667-5436>

*reversible simulation in a hyperlogic of death  
and destruction.*

— Jean Baudrillard, *Symbolic Exchange and  
Death* (2017, pp. 25–26)

## **Abstract**

In this article we analyse some of the aspects of pataphysical and conspiracy theories by examining Alex Jones's theory on frog sexual behaviour as a case study. Jones's claims possess aspects of both: they constitute an outlandish and even "wacky" theory that can sometimes be interpreted as a harmless joke, thus seeming pataphysical in nature, whilst also including aspects of conspiracy theories, such as the invocation of an alleged group of people acting in secrecy with the goal of harming the well-being of the common folk. We show that although "wacky", Jones's theory lacks the necessary aspects of pataphysical theories, and although it can be framed as conspiratorial, the main purpose of Jones's claims, which is profiting from selling expensive and ineffective products advertised by himself, means that Alex Jones' Frog Conspiracy is better understood as a fraud with conspiratorial aspects rather than a harmless conspiracy theory.

**Keywords:** 'Pataphysics; conspiracy theories; pseudoscience; endocrine-disrupting chemicals; fraud.

## Resumen

En este artículo analizamos algunos aspectos de las teorías patafísicas y conspirativas tomando como estudio de caso la teoría de Alex Jones sobre la conducta sexual de las ranas. Las afirmaciones de Jones poseen rasgos de ambas: constituyen una teoría estrafalaria e incluso “disparatada” que a veces puede interpretarse como una broma inofensiva, pareciendo así patafísica por naturaleza, al tiempo que incluyen también aspectos propios de las teorías conspirativas, como la invocación de un supuesto grupo de personas que actúa en secreto con el objetivo de perjudicar el bienestar de la gente común. Mostramos que, aunque “disparatada”, la teoría de Jones carece de los rasgos necesarios de las teorías patafísicas y que, aunque pueda enmarcarse como conspirativa, el propósito principal de las afirmaciones de Jones, que es lucrarse mediante la venta de productos caros e ineficaces que él mismo publicita, hace que la conspiración de las ranas de Alex Jones se entienda mejor como un fraude con rasgos conspirativos que como una teoría de la conspiración inofensiva.

**Palabras clave:** Patafísica; teorías de la conspiración; pseudociencia; químicos disruptores endocrinos; fraude.



## Introducción

Alex Jones, a figure known for promoting conspiracy theories, uploaded a video to YouTube in October 2015 claiming that endocrine-disrupting chemicals (EDCs) were making the frogs engage in behaviour with individuals of their same sex, and that tap water was being contaminated with the same EDCs in order to make people turn into this same behaviour and decrease the human population (Boast, 2022, pp. 661–662). Hereinafter, we shall refer to these ideas as the Frog Conspiracy (FC).

Although the FC is generally framed as a paradigmatic conspiracy theory (Boast, 2022; Dentith, 2014), some may question whether FC can *really* qualify as such, since in some instances Jones's comical delivery of the issue may lead some people into believing that it is a joke rather than a serious conspiracy accusation (Boast, 2022, p. 665). Even further, the sometimes “wacky” nature of Jones’s discourse may position FC, perhaps erroneously, as a 'pataphysical theory, in contrast to a serious conspiracy claim (cf. Hugill, 2012, p. xvi). Nonetheless, the idea of a secretive and malignant group pulling the strings, which in the case of FC takes the form of “the government” plotting against the human population, makes it harder to abandon the position of FC as a proper conspiracy theory (cf. Boudry, 2013, chap. 5).

We consider that the uncertainty of FC's conspiracy or 'pataphysical status provides an excellent opportunity to analyse it as a case study, from which clearer criteria for classifying claims as 'pataphysical, conspiratorial, or both may be inferred. Thus, the aims of this article are (i) to analyse whether the FC is a conspiracy theory or a pataphysical theory, and (ii) to provide some formal arguments that can be utilised in further inquiry into the conspiratorial or pataphysical status of other similar claims. The methodology of this article will be as follows: in § 2 we lay down the theoretical framework under which the FC will be analysed. We shall begin by listing the main characteristics and logical elements of both conspiracy and pataphysical theory, which will be applied in our analysis of the FC. In § 3 we conduct the aforementioned analysis on the FC, where it will be shown that (i) although possessing some elements of 'pataphysics, the FC cannot be considered as a pataphysical theory in light of it being concerned with general explanations and based on scientific findings, rather than focusing on exceptional cases and imaginary solutions, and that (ii) whilst conspiratorial in nature, considering the FC as *just* a conspiracy theory is a risky attitude that may trivialise the possibility that Jones's intention with the FC was to hoax his audience into buying his brand's ineffective and expensive water filters. In § 4 we conclude that an important aspect of any analysis regarding conspiratorial claims is whether the author of such claims

benefits from spreading misinformation, and in which way they may affect their believers and society in general.

## **The Concepts of Conspiracy and 'Pataphysics**

### ***What is a Conspiracy Theory?***

A conspiracy theory is, simply put, a theory involving a conspiracy. However, since some conspiracy theories may turn out to be true, belief in a conspiracy theory is not *prima facie* irrational (Dentith, 2014, p. 18), and appraisal or mistrust in a conspiracy theory should depend on the merits of the theory itself (Dentith, 2014, p. 20). The necessary and jointly sufficient criteria for a claim to be conspiratorial are:

- (1) *Conspirators condition*: positing the existence of agents with a plan,
- (2) *Secrecy condition*: the plan is being kept from the awareness of the public, and
- (3) *Goal condition*: the plan has a certain goal (Dentith, 2014, p. 23).

As previously noted, no criterion requires the theory to be false or irrational.

A subset of conspiracy theories which is of interest for epistemologists (and for the purposes of this article) is *the set of irrational conspiracy theories* (hereinafter simply referred

to as ICTs). ICTs rely on a strategy of immunisation and spurious information: they present any anomaly for their hypothesis as the result of supposed misinformation given by the conspirators to the public in order to misguide lay people into believing the conspirators and not taking into account the alleged truth. Therefore, any anomaly in the ICT can be considered as evidence for the belief system (Boudry, 2013). To further solidify mistrust in “official” sources of information, ICTs usually include various examples of people associated with “officialdom” lying to the public and spreading misinformation (Dentith, 2014, p. 138). Thus, “[n]o amount of evidence shifts these people’s beliefs, even when they are regarded as technically informed and have good reputations in other fields” (Wilkins, 2013, p. 409). It seems at first glance that ICTs are paradigmatically unlike scientific explanations, not because falsifiability alone exhausts the matter (Popper, 1962, p. 36), but because their immunising structure blocks the public criticism, evidential revision, and disciplined comparative assessment on which scientific inquiry depends. In that sense, their unfalsifiability is best understood not as an isolated defect, but as one manifestation of a broader failure of intersubjective epistemic evaluability (Contreras, 2026, p. 102). In the following paragraphs we will further analyse some models of scientific explanation and show that the aforementioned aspects of ICTs, namely the strategy of immunisation and spurious information, guarantee their failure to adapt to any such scientific model.

In order to analyse the logic (or logics) of the various ICTs, we must define what we mean by logic. If we choose definitions along the lines of “logic is the study of correct discourse”, the refutation of the argument is likely focused on the fallacious points made by the speaker. In other words, we would focus on the person rather than the facts and evidence. Science or whatever it wants to call itself has a responsibility to engage in dialogue and debate, and to counter any argument that is posited against the “scientifically proved”. Conspiracy theorists both represent and challenge the ongoing work of the so-called scientist. Therefore, a scientist cannot indulge in *ad hominem* fallacies against conspiracists, as such attitudes have always existed and always will. The best they can do is to take both the arguments (in their content) and the evidence postulated by the conspiracy theorists of the moment, and apply the respective scientific analysis to the case in question.

As mentioned in our previous paragraph, our definition of logic differs from that related to informal logic (although that does not mean that we do not use the study of argumentation in some cases). Our definition of logic will be “the study of correct reasoning” (Enderton, 2001, p. xii). Why? Much of conspiracy theorising focuses on offering a different explanation of how (or why, when, what, etc.) an event actually happened, or how certain states of affairs or facts, or even processes, really are. Therefore, the right thing

to do is to use logic to analyse the reasoning behind the explanations offered and their rationale.

To offer a scientific explanation we have multiple models, the best known of which are: the Deductive-Nomological Model (DN), Statistical Relevance Model (SR), Causal Mechanistic Model (CM), Unificationist Explanations (UE) and Pragmatic Theories of Explanation (PTE). By analysing which of all the models could be the right one for an ICT, and, if an ICT can be scientifically explained through one of these models, then such a conspiracy becomes science.

On the standard models of scientific explanation, irrational conspiracy theories (ICTs) fall short unless they are transformed into something much closer to ordinary science (Woodward & Ross, 2021). On the DN model, an explanation is a sound deductive argument whose explanans is true and essentially involves at least one law of nature, so that the explanandum is shown to be *nomically expectable*; ICTs, which rely on falsehoods and, at best, accidental generalisations, cannot meet this law-involving requirement. On Salmon's SR model, explanations must organise objectively homogeneous partitions of a population in terms of statistically relevant factors; ICTs typically treat statistically irrelevant coincidences as significant and ignore well-supported probabilistic structure, so they fail to instantiate the required relevance relations. On the CM model, explanations must trace continuous causal processes

and interactions that transmit marks or conserved quantities; posits of hidden conspirators and occult influences that cannot be connected to such processes function, at best, like the “hexed salt” example and lack explanatory import. Unificationist accounts take good explanations to be derivations that belong to an *explanatory store* of argument patterns which maximally unify the accepted corpus of belief by deriving many phenomena from a small number of stringent patterns; ICTs, which depend on highly specialised and ad hoc derivation patterns that do not increase unification, are therefore excluded from UE. Finally, on PTE accounts, explanations are answers to why-questions relative to a contrast class and a context-dependent relevance relation, but admissible answers in scientific contexts are constrained by empirically adequate background theory; ICTs that conflict with, or float free of, such theory cannot count as correct scientific explanations, so an ICT only qualifies under these models if its core claims are recast so as to satisfy their lawlike, statistical, causal-mechanical, unificatory or pragmatic constraints—which is precisely the point at which it ceases to be *irrational* in the intended sense.

Hence, ICTs are incompatible with scientific models of explanation because they systematically fail to satisfy the higher-order standards that characterise our best scientific theories: they lack robust explanatory–predictive efficacy, their empirical record does not warrant approximate ontic

adequacy, and their immunising structure undermines intersubjective epistemic evaluability.

## **What is a Pataphysical Theory?**

'Pataphysics, a term coined by Alfred Jarry, is succinctly defined as "the science of imaginary solutions", representing a deliberate parody and radical subversion of traditional scientific rationality. It emerged as a critical response to positivism, rationalism, and the scientific pretensions of complete explanatory coherence (Hugill, 2012, pp. 3–7). Unlike our best scientific theories, which aim to secure explanatory–predictive success under publicly assessable standards of criticism and revision, 'pataphysics explicitly engages with the anomalous, the exceptional, and the particular. Thus, a pataphysical theory does not concern itself with typical or regular phenomena, but rather with singularities and absurdities, treating these deviations not as errors or mere noise but as intrinsically valuable and meaningful.

To illustrate a paradigmatic pataphysical theory, consider the following passage about *the principle of variance* (anomalos), one of the pillars of 'pataphysics:

... the laws of the universe are not laws but [frequent correlations of exceptions, accidental data] ... Rules must efface the idiocracy of the *anomalos*, but



ironically such a rule about rules already risks the anomaly of paradox itself. While a metaphysical science must rule out exceptions, such exceptions are the rule (in which case they are no longer exceptions); instead, the rule is itself the exception in a 'pataphysical science that rules out the rule. The science of 'pataphysics delights in such paradoxes because its logic studies what logic exempts. (Bök, 2002, p. 39)

In contrast to scientific approaches that exclude or dismiss anomalies to reinforce the stability of theoretical frameworks, pataphysical inquiry seeks out these aberrations to reveal the inherent instability, paradox, or absurdity latent in any explanatory system. Thus, a pataphysical theory does not propose solutions in the scientific sense—as we will immediately point out—but rather imagines possibilities that subvert conventional expectations and undermine the perceived necessity of standard explanatory models.

One essential characteristic of pataphysical theories is their deliberate playful attitude towards explanation. As Alfred Jarry himself famously defined it, 'pataphysics is the “science of imaginary solutions”, intentionally producing explanations and scenarios that are surreal, whimsical, or outright absurd, and whose coherence resides primarily within the fictional or aesthetic realm rather than empirical validation. Such theories typically foreground their own

impossibility or impracticality as explanatory devices, openly acknowledging their detachment from empirical correspondence and thereby making room for literary or philosophical experimentation unconstrained by the standards that govern our best scientific theories.

This approach, however, should not be confused merely with nonsense or arbitrary fiction. 'Pataphysics involves rigorous attention to internal logical coherence and meticulous development of imaginary scenarios that critically engage with the conventions and limits of scientific rationality itself. It presents itself as a kind of philosophical satire or critical parody, but one whose ultimate goal is not merely comedic relief but philosophical provocation (Bök, 2002, p. 39; see also Hugill, 2012). By highlighting the arbitrary or contingent foundations of all theoretical systems, 'pataphysics questions the legitimacy and claims of certainty often implicit in scientific or pseudoscientific discourses, exposing their vulnerability to collapse under their own internal logic.

To put it in simple terms, the properties of 'pataphysics would be:

- (1) *Focus on the imaginary*: 'pataphysics deals with solutions and phenomena that do not necessarily have a direct correlation with observable reality,

allowing for the exploration of the absurd and the fantastic (Hugill, 2012, pp. 3–7),

(2) *Study of exceptions*: Whereas science focuses on generalities and regularities, 'pataphysics emphasizes exceptions and anomalies, considering them as offering a deeper understanding of reality (Brotchie, 1995, pp. 29–32), and

(3) *Subversion of conventional laws*: This discipline questions and subverts accepted laws and principles, proposing alternatives that challenge traditional logic (Shattuck, 1961, p. 242).

In contrast, the properties of our best scientific theories may be summarised as follows:

(1) *Explanatory–predictive efficacy*: They successfully account for the relevant phenomena and generate reliable predictions within their proper domain (Parker, 2020, p. 460),

(2) *Approximate ontic adequacy*: Their sustained empirical success is best understood as evidence that the entities, relations, and properties they posit track, at least approximately, the structure of reality (Psillos, 2005, p. 385), and

(3) *Intersubjective epistemic evaluability*: Their claims, methods, and standards of justification are publicly assessable, criticisable, and revisable by competent inquirers (Contreras, 2026, p. 102).

Moreover, pataphysical theories often deliberately blur the lines between seriousness and parody, thereby complicating straightforward interpretations and forcing audiences into a reflective stance regarding the epistemological foundations and cultural assumptions underlying conventional scientific discourse.

In light of this characterisation, 'pataphysics serves as a valuable tool for understanding theories that, though superficially absurd or bizarre, still function as critical reflections upon the epistemic practices of mainstream science. It is this critical, self-aware, and meta-theoretical dimension that distinguishes genuine pataphysical inquiry from mere nonsense, outright fraud, or irrational conspiracy theories.

This expanded and nuanced characterisation of 'pataphysics offers an excellent framework which, together with the clarifications we have made about conspiracy theories, allows us to evaluate specific cases in which an explanatory proposal contains elements that cast doubt on its logical integrity or even on whether it retains a shred of sense. In what follows, we shall rigorously evaluate Alex Jones's FC.

This examination will allow us to clearly discern whether the FC meets the sophisticated criteria of a pataphysical theory, or whether it should rather be understood primarily within the categories of conspiracy theory and fraudulent pseudoscience.

### **Alex Jones's Frog Conspiracy, a Case Study**

It should be noted that, due to instructions from the funding body for this book, we are not permitted to use the term employed by Jones in his speech, nor the neutral term used in reputable academic sources to refer to same-sex sexual behaviour. Out of courtesy, and for lack of a better term, we will refer to individuals who engage in this behaviour as individuals sexually attracted to same sex individuals, or SASSI individuals for short.

Jones's FC is significantly inspired by the research of Dr. Tyrone Hayes (see e.g. Hayes et al., 2002; Hayes et al., 2003; Hayes et al., 2010), from which Hayes concludes that atrazine, one of the most commonly used pesticides, generates sexual changes in amphibians, which makes them ““abnormal,” “gender-bending,” “[SASSI],” and therefore “unsuccessful at mating”” (O’Laughlin, 2020, p. 1). It is important to note, however, that Hayes's concern about the issue is not framed as highlighting some inherently bad or erroneous aspect of gender change or SASSIness, but rather on the decline of the population of amphibians, although it is

true that “Hayes’s creative performance style, which incorporates rhyme and comedy, has helped popularize inflammatory language about intersex frogs and make this toxic animal figure all the more compelling” (O’Laughlin, 2020, p. 2). Thus, Hayes’s discourse about his research, intentionally or not, may have caused a rhetoric of sexual alterations as pathological or bad. We shall not analyse whether Hayes’s discourse is homophobic, which may be the concern of another work. Otherwise, the conservative funding source for this book might have rejected this chapter. Instead, we shall focus on Jones’s take. As mentioned, our interest is to analyse whether Jones’s FC is a real ICT or rather a pataphysical theory. As mentioned, our interest is to analyse whether Jones’s FC is a real ICT or rather a pataphysical theory.

Jones adds to the theory the alleged fact that the government supposedly plays an active role in deliberately putting EDCs in tap water (O’Laughlin, 2020, p. 2). The objective of this intervention is to reduce human population, for if frogs are turning SASSI due to EDCs found in tap water, then humans are at risk of becoming SASSI as well, and thus, by exposure to EDCs human population would decline too (Kacala, 2018). However, Jones’s reasoning has neither a priori validity nor a posteriori support.

For instance, although humans have phenotypic traits related to gender that allow an easy observational

discrimination between male and female, phenotypic differences among frogs (e.g. *Xenopus laevis*) are more nuanced and hardly useful for the same purposes, due to the small differences between male and female frogs (O’Laughlin, 2020, pp. 10–11). Also, certain features of genetic sex in frogs warrant important contrasts with human genetic sex. In humans, genetic sex depends on the combination of X and Y chromosomes, two kinds of chromosomes which are visually distinguishable. Y is regarded as a dominant chromosome in that its presence generates ontogenetic changes to the individual so that it develops male sexual traits, thus XY individuals (generally) have a male phenotype, whilst XX individuals are (generally) phenotypically female. In *Xenopus laevis*, however, the situation is quite different: genetic sex depends on the combination of W and Z chromosomes, a pair which is visually indistinguishable and where the dominant chromosome, W, generates changes in individuals so that they develop female sexual traits. Thus, ZZ individuals are phenotypically male, whereas ZW individuals are phenotypically female (O’Laughlin, 2020, pp. 11–12). With this in mind, the conclusion that “EDCs in tap water that “turn frogs SASSI” turn humans SASSI too” is, at least a priori, not warranted.

Regarding empirical evidence, a review from the 90s (Loosli, 1995) suggests that atrazine exposure does not have clinical repercussions in humans, be it chronic or acute, although there was no explicit inquiry into possible sexual

alterations. It is also the case that research shows inconsistent findings regarding the effects on sexual traits in *Xenopus laevis*, and other studies show that the evidence on the matter regarding different species of amphibians is different—that is, that atrazine does not produce sexual alterations— or inconclusive (Gammon et al., 2005, pp. 344–346). In the same work, no mention of human sexual alterations is made. Another review (Vandenberg et al., 2012) suggests that atrazine causes sexual alterations in *Xenopus laevis* and other species of amphibians and fish, although the findings of said review are dubious<sup>99</sup> and no mention of studies regarding the effect of atrazine on human sexual traits is made. Two months after Alex Jones’s video, an executive summary from *The Endocrine Society* was published, which highlighted the absence of studies of EDCs’ impact on human sexual traits (Gore et al., 2015, p. 600). Overall, the empirical evidence regarding the effects of EDCs on human and non-human species up until the end of the year 2015 was inconclusive.

---

<sup>99</sup> It is important to mention that Dr. Hayes was a collaborator of said review. Thus, some conflict of interests may have affected the selection of studies on the review. However, it is impossible to determine whether that is the case, as there is no section on conflicts of interests in the article. Nonetheless, Rhomberg and Goodman (2012) claim that “[Vandenberg et al.] present examples as anecdotes without attempting to review all available pertinent data, selectively citing studies without evaluating most of them or examining whether their putative examples are consistent and coherent with other relevant information” (p. 130).



With the previous knowledge in mind, it can be affirmed that up to the year 2015, the date when Alex Jones posted his original video on the FC, there were neither a priori sufficient reasons nor conclusive empirical evidence to back up the claim that EDCs may turn humans SASSI. Thus, the benefit of the doubt cannot be given to Jones, as for the previously mentioned reasons we consider that his theory was not justified. It is then reasonable to attempt a classification of Jones's theory as either conspiratorial or pataphysical, which shall be the focus of the following paragraphs.

On the one hand, his deliberately funny discourse (Boast, 2022, p. 665) may incline people to believe that his theory is just an absurd pataphysical theory rather than a serious conspiratorial accusation. In fact, the more popular he has become, the more outlandish his claims have been (Van den Bulck & Hyzen, 2019, p. 8), and thus an important portion of his audience considers him just an “innocent wacko” (Van den Bulck & Hyzen, 2019, p. 12). This conclusion, however, faces two main problems. Firstly, whilst it is true that pataphysical theories seem “wacky”, “weird”, “bizarre” or any other synonym to represent a strange idea, considering 'pataphysics anything that is out of the ordinary is a mistake, as 'pataphysics is overall its own discipline with core commitments that, if absent, a theory cannot be considered pataphysical (cf. Hugill, 2012). As a consequence, classifying the FC as 'pataphysics only on the grounds of it being “outlandish” is insufficient. Secondly, the possibility of

the FC being 'pataphysics falls flat when it is considered that Jones's sources are scientific in character (although it may be disputed whether it is *good* or *bad* science), that is, based on observable data and attempting to explain generalities rather than exceptions. Since a core aspect of 'pataphysics is the exploration of imaginary ideas concerning exceptions rather than generalities, the FC cannot be considered as such.

On the other hand, the addition of a supposed government involvement in an ulterior plan to decrease human population leads to the belief that Jones's claims are conspiratorial. Also, it is true that Alex Jones backs up his suspicion of "official" information by providing examples of people in power lying to the public and spreading misinformation (Dentith, 2014, p. 138), from where he makes a fallacious generalisation to conclude that *all* "official" information is false, which is a common strategy among ICTs (Dentith, 2014, p. 137). However, although the FC can be classified as an ICT, we consider that classifying it as *just* an ICT may leave behind some important considerations about the aim of Jones's claims.

His brand, *Infowars Life*, was depicted as a brand of health-related products such as dietary supplements, brain pills, survival shields and water filters, aimed towards his

male audience (Van den Bulck & Hyzen, 2019, p. 11).<sup>100</sup> To sell his merchandise, his strategy was to attack the non-heteronormative community and boast about his own “sexual prowess” as a medium to “prove” the quality of his products (p. 9). In 2013 he changed his business model such that this became his main source of revenue (p. 10), which, in conjunction with the fact that he did not really believe in his theories (p. 12), leads us to the conclusion that the FC was just another hoax to increase the sales of his products, in particular of the aforementioned water filters. In fact, after a 2017 broadcast where Alex Jones talked about the FC, he transmitted an extensive advertisement for his water filter products, where “the most popular item, with over a thousand reviews (if they are genuine), [was] the Alexapure Pro Water Filtration System for \$269.95, containing “the same water filter that Alex Jones and the Infowars crew use everyday”” (Boast, 2022, p. 666). When all of this is taken into consideration, although it may be true that FC is in many aspects conspiratorial, classifying it as *just* an ICT ignores the fact that his claims may rather have the purpose of deceiving his target audience into buying an expensive product that is neither efficient nor effective in doing what it is purported to do, rather than generating false knowledge for the sake of *just* a political or ideological agenda. As such, FC is better

---

<sup>100</sup> Which were, in fact, just as effective as other mainstream market alternatives, although more expensive (Van den Bulck & Hyzen, 2019, p. 11).

understood as a fraud with conspiratorial aspects rather than a harmless ICT.

## **Conclusions**

The FC, in the light of our analysis, is not a pataphysical theory. Although it superficially presents absurd and comical aspects that could be mistaken for pataphysical elements, it lacks the necessary conceptual core: it is not focused on exceptional singularities, logical anomalies or imaginary solutions explicitly conceived as critical parodies of scientific logic. On the contrary, Jones attempts to give a pseudoscientific veneer to his claim by means of alleged (albeit incorrect or misrepresented) empirical evidence that generalises an effect (chemically induced SASSIness) from false premises. Therefore, it lacks the meta-theoretical and self-reflexive component essential *to its being considered* properly pataphysical.

From the perspective of conspiracy theory, the FC does present clear elements: it implies a powerful and secret group (the government) acting against the population, employs immunising strategies in the face of contrary evidence, and seeks to explain certain events through a hidden narrative. However, unlike standard conspiracy theories, the primary function of the claim may not be political or ideological, but economic and commercial: to sell ineffective products. However, this should not lead us to fall into excesses. There

are conspiracy theories that meet standards for being, in fact, reasonable or even confirmable. Thus, the rationality or irrationality of a conspiracy theory cannot be deduced by the mere fact of being “conspiratorial”, and it should be judged in light of its own epistemic merits. To condemn a theory solely because it posits hidden agents or secret intentions is to confuse form with substance, replacing epistemic evaluation with sociological suspicion. Indeed, history is replete with verified conspiracies that, before their confirmation, would have been dismissed as implausible. The proper criterion, therefore, is not whether a claim involves a conspiracy, but whether it can meet, or at least aspire to, the higher-order standards that characterise our best scientific theories: explanatory–predictive efficacy, approximate ontic adequacy, and intersubjective epistemic evaluability. When a conspiratorial hypothesis lacks explanatory reach, advances ontic commitments unsupported by sustained empirical success, and immunises itself against public criticism and revision, it ceases to function as inquiry and instead becomes self-sealing ideology. Yet, when those virtues are present, conspiratorial hypotheses may constitute legitimate, even necessary, attempts to account for complex sociopolitical phenomena. The philosophical task is thus to maintain the distinction between the conspiratorial as an epistemic form and the irrational as an epistemic failure, avoiding both naïve credulity and blanket cynicism. Therefore, we must avoid falling into the cunning use of the term “conspiracy theory” as

a synonym for fraud or trickery. Similarly, the temptation to qualify every seemingly “wacky” theory as pataphysical should be resisted, for pataphysical theories share a set of characteristics absent from everyday wacky or nonsensical discourse.

Given the above, we would strongly argue that, although it has conspiratorial features, the FC is not a classical conspiracy theory in the strict sense, let alone a pataphysical one. What emerges more clearly is that we are facing a pseudoscientific fraud with conspiratorial overtones. That is to say, Alex Jones instrumentalises conspiracy discourse to obtain economic benefits through the sale of fraudulent and ineffective products, taking advantage of the confusion generated by the ambiguity between absurd humour, pseudoscience and political paranoia. The key factor to qualify such discourses in this way lies in their practical objectives and real material implications, and not only in the formal structure of their arguments.

In short, proposals such as those of Alex Jones do not have the necessary characteristics to be registered and recognised as pataphysical proposals or as belonging to the set of non-harmful ICTs. Rather, the FC is a theoretical construct probably designed expressly for the purpose of duping the unwary. Such schemes inevitably imply an asymmetry in the results, where the promoting agent benefits or, in any case, is not affected by any negative consequences,

but the manipulations are always harmed (to a greater or lesser extent).

Hence, it is no trivial matter to develop the theoretical tools needed to identify socially pernicious and harmful proposals. To limit oneself to mocking and ridiculing them only leaves the door open to their free diffusion in the collective imagination. The crucial question is: for how long can we go on doing this without consequences, so long as it does not affect us and the victim is always someone else?

## References

- Boast, H. (2022). Theorizing the gay frog. *Environmental Humanities*, 14(3), 661–679.  
<https://doi.org/10.1215/22011919-9962959>
- Bök, C. (2002). *'Pataphysics: The poetics of an imaginary science*. Northwestern University Press.
- Boudry, M. (2013). Loki's wager and Laudan's error: On genuine and territorial demarcation. In M. Pigliucci & M. Boudry (Eds.), *Philosophy of pseudoscience* (pp. 79–98). University of Chicago Press.  
<https://doi.org/10.7208/chicago/9780226051826.003.0006>
- Brotchie, A. (1995). *A book of surrealist games*. Shambhala Publications.
- Contreras, C. (2026). Un naturalismo minimalista. *Revista de Filosofía Homónima*, 1(1), 79–111.

<https://revistahomonima.com/2026/01/30/un-naturalismo-minimalista-constantino-contreras-hernandez/>

Dentith, M. R. (2014). *The philosophy of conspiracy theories*.  
Palgrave Macmillan.

<https://doi.org/10.1057/9781137363169>

Enderton, H. B. (2001). *A mathematical introduction to logic*  
(2nd ed.). Academic Press.

Gammon, D. W., Aldous, C. N., Carr, W. C., Jr., Sanborn, J. R., & Pfeifer, K. F. (2005). A risk assessment of atrazine use in California: Human health and ecological aspects. *Pest Management Science*, 61(4), 331–355.

<https://doi.org/10.1002/ps.1000>

Gore, A. C., Chappell, V. A., Fenton, S. E., Flaws, J. A., Nadal, A., Prins, G. S., Toppari, J., & Zoeller, R. T. (2015). Executive summary to EDC-2: The Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocrine Reviews*, 36(6), 593–602.

<https://doi.org/10.1210/er.2015-1093>

Hayes, T., Haston, K., Tsui, M., Hoang, A., Haeffele, C., & Vonk, A. (2003). Atrazine-induced hermaphroditism at 0.1 ppb in American leopard frogs (*Rana pipiens*): Laboratory and field evidence. *Environmental Health Perspectives*, 111(4), 568–575.

<https://doi.org/10.1289/ehp.5932>

Hayes, T. B., Collins, A., Lee, M., Mendoza, M., Noriega, N., Stuart, A. A., & Vonk, A. (2002). Hermaphroditic,



demasculinized frogs after exposure to the herbicide atrazine at low ecologically relevant doses. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), 5476–5480. <https://doi.org/10.1073/pnas.082121499>

Hayes, T. B., Khoury, V., Narayan, A., Nazir, M., Park, A., Brown, T., Adame, L., Chan, E., Buchholz, D., Stueve, T., & Gallipeau, S. (2010). Atrazine induces complete feminization and chemical castration in male African clawed frogs (*Xenopus laevis*). *Proceedings of the National Academy of Sciences of the United States of America*, 107(10), 4612–4617. <https://doi.org/10.1073/pnas.0909519107>

Hugill, A. (2012). *Pataphysics: A useless guide*. MIT Press.

Kacala, A. (2018, August 7). Infowars' Alex Jones has a long history of inflammatory, anti-LGBTQ speech. *NBC News*. <https://www.nbcnews.com/feature/nbc-out/infowars-alex-jones-has-long-history-inflammatory-anti-lgbtq-speech-n898431>

Loosli, R. (1995). Epidemiology of atrazine. In G. W. Ware (Ed.), *Reviews of environmental contamination and toxicology: Continuation of residue reviews* (Vol. 143, pp. 47–57). Springer. [https://doi.org/10.1007/978-1-4612-2542-3\\_2](https://doi.org/10.1007/978-1-4612-2542-3_2)

O'Laughlin, L. N. (2020). Troubling figures: Endocrine disruptors, intersex frogs, and the logics of environmental science. *Catalyst: Feminism, Theory*,

and *Technoscience*, 6(1), 1–28.  
<https://doi.org/10.28968/cftt.v6i1.32350>

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, 87(3), 457–477.  
<https://doi.org/10.1086/708691>

Popper, K. R. (1962). *Conjectures and refutations: The growth of scientific knowledge*. Basic Books.

Psillos, S. (2005). Scientific realism and metaphysics. *Ratio*, 18(4), 385–404. <https://doi.org/10.1111/j.1467-9329.2005.00301.x>

Shattuck, R. (1961). *The banquet years: The origins of the avant-garde in France, 1885 to World War I*. Anchor Books.

Van den Bulck, H., & Hyzen, A. (2019). Of lizards and ideological entrepreneurs: Alex Jones and Infowars in the relationship between populist nationalism and the post-global media ecology. *International Communication Gazette*, 82(1), 42–59.  
<https://doi.org/10.1177/1748048519880726>

Vandenberg, L. N., Colborn, T., Hayes, T. B., Heindel, J. J., Jacobs, D. R., Jr., Lee, D. H., Shioda, T., Soto, A. M., vom Saal, F. S., Welshons, W. V., Zoeller, R. T., & Myers, J. P. (2012). Hormones and endocrine-disrupting chemicals: Low-dose effects and nonmonotonic dose responses. *Endocrine Reviews*, 33(3), 378–455. <https://doi.org/10.1210/er.2011-1050>

- Wilkins, J. S. (2013). The Salem region: Two mindsets about science. In M. Pigliucci & M. Boudry (Eds.), *Philosophy of pseudoscience* (pp. 397–416). University of Chicago Press. <https://doi.org/10.7208/9780226051826-022>
- Woodward, J., & Ross, L. (2021, May 10). Scientific explanation. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/scientific-explanation>

# El auge de las revistas de baja calidad en el ámbito académico: causas, consecuencias y desafíos éticos

*The Rise of Low-Quality Journals in Academia: Causes, Consequences, and Ethical Challenges*

**Héctor Javier Bendezú Jiménez<sup>101</sup>**

## Resumen

En los últimos años, la publicación en revistas indexadas de calidad cuestionable ha crecido, lo que supone un riesgo para la integridad académica. Estas revistas, que eventualmente son descatalogadas por incumplir estándares científicos, afectan la reputación de la comunidad

---

<sup>101</sup> Universidad Nacional Mayor de San Marcos. <https://orcid.org/0000-0001-9530-6472>

investigadora —especialmente de los menos experimentados— y de sus instituciones. Este capítulo tiene como objetivo analizar las causas, consecuencias y desafíos éticos asociados al auge de las revistas de baja calidad en el ámbito académico. Para ello, se adopta un enfoque cualitativo de análisis crítico, basado en la revisión de investigaciones previas, documentos técnicos y estudios bibliométricos relevantes. Los hallazgos indican que la vinculación con editoriales de dudosa calidad responde, por un lado, a un sistema académico principalmente estructurado bajo la lógica del capitalismo académico, que incentiva la productividad cuantitativa sin normativas rigurosas en materia editorial; y por otro, a la inexperiencia o a la presión por publicar rápidamente la mayor cantidad posible de artículos. Esta combinación ha facilitado la proliferación de publicaciones de baja calidad. Se concluye que preservar la calidad académica requiere reforzar la ética científica, mejorar los mecanismos de control editorial y evaluación, y fomentar una formación investigadora crítica y comprometida con la integridad académica.

**Palabras clave:** Capitalismo académico; integridad científica; ética de investigación; revistas depredadoras; revistas indexadas.

## **Abstract**

In recent years, the publication of articles in indexed journals of questionable quality has increased, posing a risk to academic integrity. These journals, which are eventually delisted for failing to meet scientific standards, damage the reputation of the research community—particularly early-career researchers—and their institutions. This article aims to analyze the causes, consequences, and ethical challenges associated with the rise of low-quality academic journals. To this end, a qualitative approach based on critical analysis is adopted, relying on a review of previous research, technical documents, and relevant bibliometric studies. The findings indicate that engagement with dubious publishers stems, on the one hand, from an academic system shaped by the logic of academic capitalism, which incentivizes quantitative productivity without rigorous editorial standards; and on the other hand, from researchers' inexperience or pressure to publish as many articles as possible quickly. This combination has facilitated the proliferation of low-quality publications. It is concluded that safeguarding academic quality requires strengthening scientific ethics, improving editorial and peer-review control mechanisms, and fostering a research culture committed to academic integrity.

**Keywords:** Academic capitalism; scientific integrity; research; predatory journals; indexed journals.

## Introducción

A finales de la década de los 90, Slaughter y Leslie (2001) plantearon una reflexión que luego fue ampliada por Slaughter y Rhoades (2004) sobre los cambios en la educación superior, especialmente en el mundo anglosajón. Estas autoras identificaron una creciente tendencia hacia la lógica del mercado, caracterizada por la reducción de costos y la maximización de ganancias, lo que ha distorsionado la misión tradicional de las instituciones educativas: generar y compartir conocimiento, así como inculcar valores morales en los futuros profesionales. Este nuevo paradigma fue denominado *Academic Capitalism* (AC), cuya recepción y adaptación en América Latina como perspectiva de estudio aún no ha sido suficientemente explorada (Brunner et al., 2019).

En este contexto, Jessop (2018) describe cinco fases a través de las cuales la academia puede experimentar un proceso de mercantilización, comercialización y capitalización, hasta transformarse en instituciones cuyo único propósito sea la rentabilidad y el beneficio económico. Este fenómeno se observa en varios países de América Latina, donde prolifera la creación de universidades privadas masivas y de bajo costo. Por otro lado, en el ámbito de la investigación, la implementación de incentivos ha impulsado un aumento en la producción científica. Sin embargo, este modelo también ha dado lugar a una proliferación de investigaciones masivas

con escasa rigurosidad científica, lo que plantea desafíos para la calidad y credibilidad del conocimiento generado en estos entornos académicos.

Paralelamente, a finales del siglo XX surgió el Movimiento de Acceso Abierto (Open Access), cuyo objetivo era democratizar el acceso a los descubrimientos científicos, trasladando los costos de publicación a los autores. Sin embargo, esta práctica fue posteriormente explotada por algunas editoriales con fines lucrativos, en detrimento de la calidad de las investigaciones (Bertoglia & Águila, 2018). Si bien esta tendencia parecía, en teoría, una respuesta al *Academic Capitalism*, con el tiempo fue instrumentalizada por las propias universidades para consolidar su burocracia y reforzar la lógica mercantilista. Como resultado, se debilitó el modelo tradicional de investigación, que priorizaba la colaboración y la libre distribución del conocimiento sin fines de lucro (Hostler, 2024).

Saura y Caballero (2021) expresan su preocupación respecto al culto a la metrificación en la producción de conocimiento. Aunque esta tendencia se originó a finales del siglo pasado, ha sido durante el auge del capitalismo digital, la globalización y el uso de herramientas de *Big Data* e inteligencia artificial cuando se ha logrado cuantificar a nivel mundial a través de diversos índices de impacto. Entre estos se incluyen rankings universitarios como el *Academic Ranking of World Universities*, el *Times Higher Education*



*World University Ranking* y el *QS World Rankings*, así como métricas para la comunidad investigadora, tales como el *Science Citation Index*, el *SCImago Journal Rank*, el *Google Scholar H-Index*, *Altmetrics* y el *ResearchGate Score*, entre otros.

Si bien la competencia internacional no es inherentemente negativa, esta dinámica puede inducir a las instituciones públicas y privadas dedicadas a la generación de conocimiento, así como a su personal investigador, a optar por estrategias que garanticen una alta puntuación en estos indicadores. Esto se traduce en una priorización de la producción cuantitativa sobre la calidad académica, con el objetivo de atraer más financiamiento público y privado y mejorar su posición en las clasificaciones internacionales. En consecuencia, se consolida un modelo en el que la cantidad de publicaciones se equipara erróneamente con la excelencia académica.

Este contexto ha propiciado el surgimiento de las denominadas revistas depredadoras, dado que la constante publicación de artículos científicos en revistas de alto impacto se ha convertido en una necesidad para los docentes, ya que les permite acceder a diversos beneficios académicos, profesionales y económicos (Jiménez-Yañez & Colmenares-Díaz, 2022). Kendall (2021) advierte que este tipo de decisiones no pueden entenderse como simples actos individuales, sino como respuestas funcionales a un sistema

de incentivos que privilegia la visibilidad y el conteo de publicaciones por encima del compromiso con la calidad científica. Como consecuencia, muchos profesionales de la investigación optan por evitar el esfuerzo y la rigurosidad que implica publicar por la vía adecuada y, en su lugar, buscan atajos para agilizar el proceso.

Esta demanda ha generado un mercado de editoriales que garantizan publicaciones rápidas y con estándares de calidad laxos, lo que ha permitido, de forma alarmante, que muchas de estas revistas ingresen —aunque sea temporalmente— a bases de datos reconocidas, poniendo en entredicho los actuales mecanismos de validación editorial (Koerber et al., 2020).

En este sentido, Mills e Inouye (2021), a partir de una revisión sistemática de 16 estudios sobre las motivaciones de personas investigadoras de Asia, África, Sudamérica, Oriente Medio y otras regiones, identifican como principales razones para recurrir a editoriales depredadoras los incentivos monetarios, las oportunidades de ascenso institucional —como promociones o mejoras laborales— y los requisitos exigidos para postular a programas de maestría o doctorado.

En esta línea, Mertkan et al. (2021) destacan el creciente interés de la comunidad académica en estudiar el fenómeno de las revistas depredadoras, dada su relevancia en el ámbito científico. Entre estas investigaciones se encuentran

diversos estudios bibliométricos que analizan revistas potencialmente depredadoras, muchas de las cuales han sido clasificadas como discontinuadas tras haber estado alojadas temporalmente en bases de datos de prestigio. En suma, destacan los estudios de Cortegiani et al. (2020), Marina y Sterligov (2021) y Macháček y Srholec (2021). No obstante, este último fue objeto de retractación por parte de la editora jefa debido a problemas metodológicos (Macháček & Srholec, 2022).

En este escenario, el análisis de este fenómeno resulta crucial, ya que pone en evidencia las tensiones entre la creciente presión por publicar y la integridad académica, además de los efectos que este tipo de prácticas pueden tener sobre la credibilidad del conocimiento generado y el prestigio de las instituciones de educación superior.

En este sentido, surge una cuestión fundamental que merece ser analizada en profundidad: *¿Cuál es el impacto real de publicar en este tipo de revistas para los investigadores?*

El efecto de estas publicaciones puede abordarse desde múltiples dimensiones. En el corto plazo, la inclusión de artículos en revistas indexadas, aunque de baja calidad, puede brindar beneficios inmediatos como el cumplimiento de requisitos para becas, programas de posgrado o ascensos. Sin embargo, *¿cuál es su impacto a mediano y largo plazo? ¿Esta*

*práctica conlleva riesgos significativos tanto para los autores como para la comunidad académica en general?*

*Desde una perspectiva individual, cabe preguntarse: ¿puede afectar la reputación de quienes publican una vez que estas revistas son desindexadas o desacreditadas? ¿Limita sus oportunidades de colaboración con instituciones de prestigio? Y a nivel institucional, ¿La proliferación de este tipo de publicaciones contribuye a la deslegitimación de los sistemas de evaluación científica y debilita el prestigio de las universidades y centros de investigación?*

Desde la perspectiva de la epistemología social, la discusión sobre las revistas depredadoras se ha vuelto relevante porque analiza cómo los factores sociales e institucionales influyen en la producción, validación y difusión del conocimiento científico. En este enfoque, el conocimiento se genera dentro de comunidades epistémicas organizadas, donde universidades, sistemas de indexación y revistas académicas cumplen un papel central en la evaluación y legitimación de los resultados científicos. Las revistas legítimas forman parte de la infraestructura del sistema de conocimiento, ya que el proceso de revisión por pares actúa como un mecanismo de control de calidad que busca asegurar rigor metodológico y contribuciones relevantes. No obstante, las revistas depredadoras, que simulan procesos editoriales legítimos y priorizan el cobro de tarifas sin una evaluación rigurosa, introducen distorsiones

en este sistema de validación. Desde esta perspectiva, este fenómeno puede interpretarse como una falla institucional en los mecanismos colectivos de certificación del conocimiento, ya que debilita la credibilidad del proceso de revisión por pares y facilita la circulación de investigaciones con escaso control académico. Además, su expansión está vinculada a incentivos estructurales del sistema académico contemporáneo, especialmente a la lógica del *publish or perish*, que presiona a los investigadores a incrementar su productividad científica.

En consecuencia, es crucial reflexionar sobre la responsabilidad de las instituciones académicas, las agencias de financiamiento y el personal investigador en la consolidación de un sistema que valore la calidad por encima de la cantidad, promoviendo estándares éticos que salvaguarden la integridad del conocimiento.

Estas interrogantes conducen al objetivo central de este capítulo: analizar las causas, consecuencias y desafíos éticos que explican el auge de las revistas de baja calidad en el ámbito académico, así como su impacto en la credibilidad de la ciencia, la reputación investigadora y la legitimidad institucional.

## **Método**

Este estudio adopta un enfoque cualitativo crítico, que permite analizar el fenómeno desde una perspectiva holística y estructural, atendiendo a la complejidad de las dinámicas académicas contemporáneas. A diferencia de los estudios basados en la cuantificación de variables aisladas, este enfoque favorece la comprensión de las interacciones y dinámicas dentro del objeto de estudio. En este caso, la academia constituye el espacio clave donde se desarrolla el fenómeno analizado, por lo que la interacción de la persona responsable de la investigación con este entorno resulta esencial para la construcción del conocimiento y la formulación de conclusiones fundamentadas (Bernal, 2016).

El análisis se basa en una revisión documental e interpretativa de literatura previa, estudios bibliométricos y normativas sobre evaluación científica. Sin embargo, más allá de una simple recopilación de literatura, se adopta un enfoque interpretativo que permite cuestionar y contextualizar los hallazgos de estos estudios a la luz del problema de investigación.

Este proceso implica identificar patrones, tensiones y vacíos en la producción académica, así como examinar el impacto de las dinámicas institucionales y estructurales en la proliferación de publicaciones en revistas de baja calidad. En este sentido, la investigación no solo busca describir el

fenómeno, sino también aportar una mirada crítica que contribuya al debate sobre la calidad y la integridad del conocimiento científico en el contexto actual (Departamento Administrativo de Ciencia, Tecnología e Innovación, 2014).

## **El ecosistema de las revistas científicas globalizadas**

En el contexto de la producción científica global, existen miles, e incluso millones, de revistas o journals que forman parte de un ecosistema académico interconectado y cada vez más influenciado por la globalización. Dentro de esta vasta red de publicaciones, es posible identificar una gran diversidad de tipos y categorías de revistas, aunque en términos generales pueden agruparse en dos grandes clases: (i) revistas científicas legítimas y (ii) revistas depredadoras.

Un error común al clasificar estas publicaciones es basarse exclusivamente en su indexación en bases de datos reconocidas, asumiendo que la presencia en estos repertorios garantiza su calidad y rigor científico. Sin embargo, esta suposición es incorrecta, ya que muchas revistas depredadoras logran obtener una indexación temporal en ciertas bases de datos, aprovechando vacíos en los criterios de selección o procesos de curaduría deficientes. Esto genera una falsa percepción de legitimidad y dificulta la distinción entre revistas de alta calidad y aquellas que operan con estándares laxos o prácticas poco éticas.

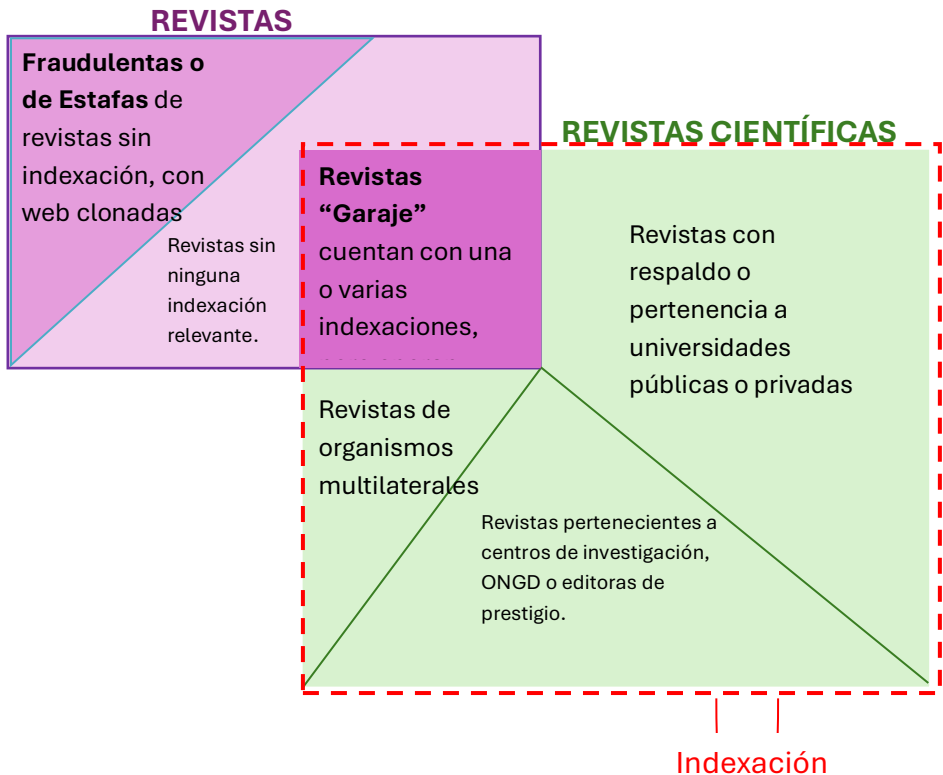
Por ello, la clasificación de las revistas debe considerar otros factores clave, como el respaldo institucional, la solidez del proceso de revisión por pares, la transparencia en los criterios editoriales, la reputación del comité editorial y la coherencia con las buenas prácticas en la comunicación científica. Solo así es posible diferenciar entre publicaciones que contribuyen genuinamente al avance del conocimiento y aquellas que buscan capitalizar la creciente presión por publicar sin garantizar la integridad académica.

En el siguiente esquema se puede apreciar, la subclasificación que existe dentro de las revistas legítimamente científicas y las depredadoras.



**Figura 1**

*Ecosistema de las revistas a nivel global*



*Nota.* Se busca representar la subclasificación existente dentro de las revistas científicas legítimas y las depredadoras. Elaboración del autor.

En este ecosistema, la indexación en bases de datos académicas cumple un rol estratégico, cuya relevancia y evolución se abordarán en el siguiente apartado.

### *La indexación en bases de datos y su rol en el ecosistema científico*

Las bases de datos bibliográficas académicas son fundamentales para organizar y difundir el conocimiento científico. Según Feather (2003), estas bases constituyen *colecciones estructuradas de información* que incluyen literatura científica en diversas formas, como artículos de revistas, ponencias, informes y otros documentos académicos. Su principal característica es la existencia de *catálogos en línea* que permiten el almacenamiento y la recuperación de información mediante sistemas de indexación por materias.

El origen de estas bases de datos responde a la necesidad de la academia de *organizar y facilitar el acceso a la producción científica*, un reto que en sus inicios se resolvía con registros bibliográficos físicos. Sin embargo, con el desarrollo de las tecnologías digitales y, en particular, con la aparición de Internet, se construyeron *infraestructuras electrónicas* que optimizaron el acceso a la información. A partir de la década de 1970, comenzaron a surgir servicios que ofrecían *bases de datos académicas en línea*, posibilitando búsquedas más avanzadas por *palabras clave, autor, título y*

*otros criterios* (Borgman, 2007). Aunque en sus primeras etapas estos sistemas eran complejos y costosos para los usuarios, su evolución ha permitido que, en la actualidad, sean herramientas esenciales para la investigación científica.

Entre las diversas bases de datos existentes, *Web of Science (WoS)*, *Scopus* y *Scielo* se destacan por su *rigurosidad en la selección de publicaciones* y por la utilización de métricas avanzadas para evaluar el impacto de la producción científica.

**Web of Science (WoS): pionera en la indexación académica.** Fundada en 1997, es una de las más prestigiosas por sus estrictos criterios de inclusión. Ofrece métricas como el índice de impacto y el índice H, permitiendo visualizar redes de citas y evaluar el alcance de las investigaciones (Birkle et al., 2020).

**Scopus: rigor y evaluación continua.** Desarrollada por Elsevier, cuenta con un consejo de expertos que supervisa los criterios de indexación. Utiliza el índice H y el SJR (SCImago Journal Rank) como herramientas para medir impacto y relevancia, y realiza revisiones periódicas para asegurar la calidad (Baas et al., 2020).

**SciELO: accesibilidad y ciencia abierta.** Iniciativa latinoamericana nacida en 1998 para visibilizar la ciencia de países en desarrollo. Se centra en el acceso abierto,

incluye revisión por pares y ofrece el SciELO Citation Index como métrica de impacto. Tiene un enfoque fuerte en ciencias sociales y salud pública.

Tanto Web of Science como Scopus han consolidado su prestigio como referentes en la *indexación de revistas científicas de alto impacto*. Su función no solo es *organizar el conocimiento*, sino también garantizar que las publicaciones científicas cumplan con *criterios de calidad*, evitando la proliferación de revistas de dudosa credibilidad. Sin embargo, la existencia de *revistas de baja calidad dentro de estas bases de datos*, conocidas como "*revistas garaje*", demuestra que, aunque los filtros son rigurosos, *no son infalibles*. Por ello, la comunidad científica debe mantenerse alerta y fomentar prácticas de publicación *responsables y éticas*, asegurando que el conocimiento generado contribuya realmente al avance de la ciencia.

### *Las revistas científicas legítimas*

El presente grupo (Figura 1), marcado en *color verde*, es de mayor cantidad y a diferencia de las revistas depredadoras y las denominadas "revistas garaje", se distinguen por mantener *altos estándares de calidad* respaldados por *métricas e indicadores bibliométricos*. Estas publicaciones son reconocidas por estar alojadas en *bases de datos académicas de prestigio*, como *Scopus, Web of Science y SciELO*, entre otras.

Las revistas indexadas presentan una serie de características clave que garantizan su *rigurosidad y credibilidad en el ámbito académico*:

- **Respaldo institucional:** Pertenecen o están asociadas a universidades, centros de investigación o entidades académicas y científicas reconocidas.
- **Evaluación por pares:** Implementan un sistema de revisión doble ciego o abierto, en el que expertos en la materia analizan y validan los artículos antes de su publicación.
- **Periodicidad establecida:** Suelen publicar de manera semestral, trimestral o anual, evitando la sobreproducción de contenidos sin control de calidad.
- **Indexación en bases de datos reconocidas:** Forman parte de sistemas de indización confiables que certifican su calidad y les otorgan mayor visibilidad en la comunidad científica.
- **Normativas y códigos de ética:** Cuentan con lineamientos editoriales claros, políticas de integridad académica y adhesión a principios éticos internacionales en la publicación científica.

- **Identificadores digitales (DOI):** Asignan a cada artículo un Digital Object Identifier (DOI), lo que facilita su trazabilidad y garantiza su permanencia en el tiempo.
- **Tasa de rechazo elevada:** Mantienen criterios estrictos de selección, lo que implica que solo un porcentaje reducido de los manuscritos enviados logran ser publicados tras un riguroso proceso de evaluación.
- **Costos de publicación:** Dependiendo de la política de la revista, algunas exigen el pago de tasas de publicación, especialmente en el caso de los artículos de acceso abierto (Open Access).

Estas revistas desempeñan un papel fundamental en la *validación y diseminación del conocimiento científico*, ya que garantizan que los estudios publicados cumplen con *criterios de originalidad, relevancia y rigor metodológico*. Su impacto en la comunidad académica es significativo, pues no solo aseguran la calidad de la producción científica, sino que también sirven como *referentes para la evaluación del desempeño investigador*, influyendo en procesos de promoción profesional y en la asignación de financiamiento para proyectos de investigación (Jiménez-Yañez & Colmenares-Díaz, 2022).

No obstante, el crecimiento acelerado de la industria editorial académica y la *presión por publicar* han llevado a

que algunas revistas indexadas adopten *prácticas cuestionables*, como la proliferación de publicaciones excesivas en períodos cortos o la flexibilización de los criterios de evaluación. Por ello, la comunidad científica debe *fortalecer los mecanismos de control de calidad*, garantizando que la indexación continúe siendo un *sello de excelencia y no solo un requisito burocrático* en el sistema académico.

### **Clasificación de revistas científicas legítimas.**

Al analizar con mayor profundidad el grupo de las revistas científicas legítimas, es posible establecer una clasificación que permita diferenciar su naturaleza, estructura y funcionamiento. La siguiente tipología, basada en la experiencia del autor y en el análisis de criterios de calidad editorial y académica, permite una mejor comprensión del ecosistema de publicaciones científicas.

- ***Revistas académicas respaldadas por universidades o centros de investigación reconocidos***

Este tipo de revistas es publicado por *universidades (públicas o privadas) o centros de investigación altamente prestigiosos*. Se caracterizan por:

- *Contar con un respaldo institucional sólido y verificable*, lo que garantiza su credibilidad.

- *Tener una alta demanda de publicación*, lo que conlleva a *tasas de rechazo elevadas* (generalmente superiores al 70%).
- *Publicar con periodicidad semestral o anual*, favoreciendo una mayor rigurosidad en la selección de artículos.
- *No permitir publicaciones continuas de una misma autoría dentro de un año*, para evitar la monopolización del contenido por un reducido grupo de firmantes.
- *No exigir pagos por procesamiento o publicación de artículos* (*Article Processing Charges – APC*), lo que evita sesgos económicos en la selección de manuscritos.

Dado su *riguroso proceso de revisión por pares y altos estándares de calidad*, estas revistas son altamente valoradas en la comunidad científica y suelen figurar en las primeras posiciones de rankings internacionales de impacto académico.

- ***Revistas pertenecientes a organismos internacionales y entidades estatales***

Si bien son menos frecuentes que las revistas académicas universitarias, este grupo de publicaciones científicas es



gestionado por *organismos multilaterales, agencias gubernamentales y entidades estatales de investigación*. Sus principales características incluyen:

- *Un alto grado de competencia*, dado que suelen recibir contribuciones de profesionales de la investigación con trayectorias consolidadas.
- *Plazos de publicación prolongados*, similares a los de las revistas académicas universitarias, debido a la rigurosidad de su proceso editorial.
- *Acceso abierto en la mayoría de los casos*, ya que muchas de estas publicaciones están financiadas con fondos públicos o por cooperación internacional.

Dado su enfoque institucional, estas revistas suelen especializarse en áreas estratégicas como *políticas públicas, desarrollo sostenible, salud, educación, tecnología, economía y ciencias sociales*, buscando generar impacto en la toma de decisiones gubernamentales y en la formulación de políticas basadas en evidencia.

- ***Revistas pertenecientes a centros de investigación, ONG y editoriales de prestigio***

Un tercer grupo de revistas científicas legítimas está compuesto por publicaciones gestionadas por *centros de*

*investigación independientes, organizaciones no gubernamentales para el desarrollo (ONGD) y editoriales científicas reconocidas.* Estas revistas presentan las siguientes características:

- *Pertenecer a entidades con un prestigio consolidado,* aunque sin el mismo nivel de respaldo institucional que las revistas universitarias o estatales.
- *Contar con financiamiento mixto,* combinando recursos privados, subvenciones y en algunos casos modelos de pago por publicación (APC), aunque con criterios éticos claros.
- *Publicar con una frecuencia más flexible,* pudiendo ser trimestrales, cuatrimestrales o semestrales.
- *Tener un proceso de evaluación por pares riguroso,* aunque en algunos casos presentan tasas de rechazo ligeramente menores en comparación con revistas universitarias.

Muchas de estas revistas se encuentran indexadas en bases de datos reconocidas como *Scopus*, *Web of Science* y *Scielo*, y tienen un impacto significativo en sus respectivas áreas de especialización. Algunas editoriales de prestigio en este ámbito incluyen *Elsevier*, *Springer*, *Wiley* y *Taylor & Francis*, entre otras.

En líneas generales la existencia de diferentes tipos de revistas científicas legítimas refleja la diversidad del ecosistema académico y editorial. Si bien todas las revistas mencionadas cumplen con estándares de calidad, es fundamental que los investigadores elijan cuidadosamente dónde publicar sus trabajos, considerando no solo la indexación, sino también el *prestigio institucional*, el *rigor del proceso de revisión* y el *modelo de acceso*. Esta selección no solo impacta en la visibilidad y credibilidad de la investigación, sino también en el desarrollo del conocimiento en sus respectivas disciplinas.

En los últimos años, la publicación de artículos, manuscritos o *papers* de investigación se ha convertido en un *requisito obligatorio* para el avance profesional y académico de investigadores, docentes y estudiantes de posgrado. Esta exigencia ha fomentado una *competencia académica intensa*, en la que los creadores de conocimiento buscan publicar en revistas científicas de prestigio, sometiéndose a *rigurosos procesos de evaluación*.

### *El proceso de publicación en las revistas científicas*

El proceso de publicación en revistas indexadas implica superar dos niveles de revisión: una *evaluación editorial inicial*, en la que se verifica la adecuación temática y formal del manuscrito, y una *revisión por pares externos* bajo

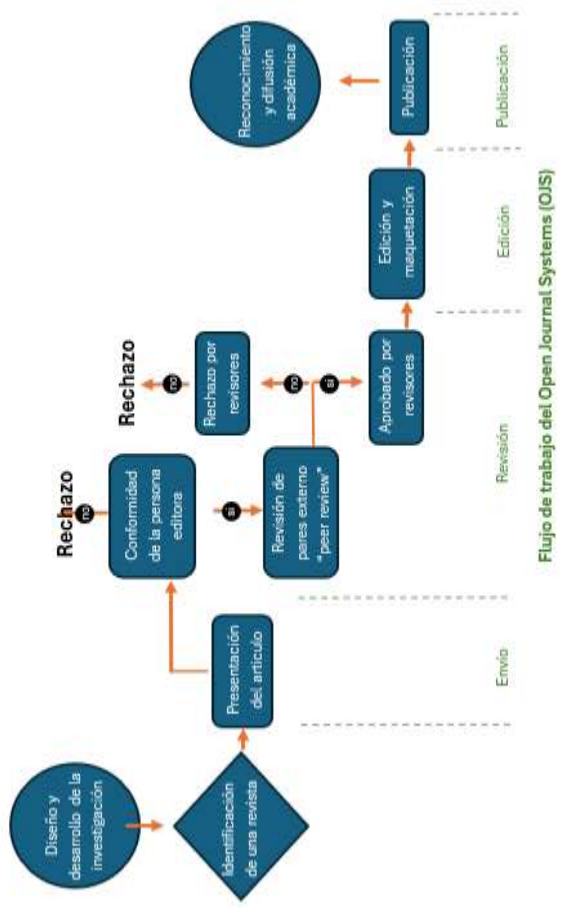
el sistema *doble ciego*, donde se evalúa la calidad, originalidad y rigor metodológico del estudio.

*La tasa de aceptación en muchas revistas de alto impacto es inferior al 20%*, lo que refleja el nivel de exigencia de estas publicaciones y la dificultad para lograr la difusión de investigaciones en medios de calidad. A esto se suma el *tiempo requerido para la evaluación y aprobación de un artículo*, que en el mejor de los casos oscila entre *8 y 10 meses*, aunque en revistas con una alta demanda de publicaciones puede extenderse hasta *un año o más*.

Sin embargo, más allá de los desafíos que implica este proceso, el *objetivo central* de la publicación científica debe ser *la generación de conocimiento ético, riguroso y transparente*, asegurando que los hallazgos de las investigaciones sean *accesibles para la comunidad académica y la sociedad en general*. La presión por publicar, aunque fomenta el crecimiento del conocimiento, también ha dado lugar a *prácticas cuestionables*, como la búsqueda de atajos mediante revistas de baja calidad.

**Figura 2**

*Flujograma para publicar manuscritos en revistas indexadas*



**Nota.** El flujograma refleja los pasos que se desarrollan al presentar un artículo a una revista indexada. Elaboración del autor.

A partir de la Figura 2, se puede observar que, si bien el proceso de publicación de un artículo científico no es excesivamente complejo en términos procedimentales, sí requiere un alto nivel de calidad académica, esfuerzo, dedicación y una considerable inversión de tiempo. Además, implica un fuerte compromiso con la integridad científica, dado que el camino hacia la excelencia académica no solo supone superar rigurosos procesos de revisión por pares, sino también afrontar posibles rechazos, revisiones extensas y múltiples rondas de corrección.

En este sentido, uno de los mayores desafíos para los proponentes de un manuscrito es *mantener la resiliencia y la ética profesional*, evitando caer en la desmotivación o en la tentación de recurrir a atajos que comprometan la calidad y credibilidad de sus publicaciones. La presión por publicar — ya sea para obtener reconocimiento, acceder a incentivos académicos o mejorar la posición profesional— puede llevar a algunos académicos a optar por alternativas cuestionables, como las revistas depredadoras, que prometen publicaciones rápidas y con mínimos requisitos de calidad.

Por ello, más allá de comprender el procedimiento de publicación, es crucial *fomentar una cultura de ética*,

*paciencia y rigor académico*, en la que los investigadores valoren el proceso tanto como el resultado. La verdadera excelencia académica no radica en la cantidad de publicaciones, sino en su impacto y contribución real al avance del conocimiento.

Por el contrario, en un intento por eludir el riguroso proceso de publicación en revistas científicas de alta calidad, algunas profesionales y académicos han optado por estrategias cuestionables que priorizan la acumulación de publicaciones sobre la relevancia y el impacto real de sus contribuciones. Entre estas prácticas, destacan las llamadas "*granjas de científicos*", donde profesionales y docentes de ética dudosa pagan para aparecer como firmantes en artículos de temas ajenos a su especialidad, abordando problemas de países que nunca han visitado y sin haber participado activamente en la investigación.

Este fenómeno se agrava debido a la presión por publicar en revistas indexadas en bases de datos reconocidas como *Scielo*, *Scopus* y *Web of Science (WoS)*, las cuales son consideradas referencia en muchos sistemas de evaluación académica. Por ejemplo, en Perú, el *Consejo Nacional de Ciencia, Tecnología e Innovación (Concytec)* solo reconoce artículos publicados en estas bases de datos para la calificación y clasificación de investigadores dentro del *Registro Nacional Científico, Tecnológico y de Innovación Tecnológica (Renacyt)*. Esta situación ha generado un

incentivo perverso, donde la prioridad ya no es la generación de conocimiento, sino la simple obtención de puntuaciones que permitan acceder a nombramientos, promociones o ratificaciones docentes.

Sin embargo, incluso estas bases de datos no están exentas de fallos. A pesar de sus sistemas de control, *existen revistas de baja calidad o "de garaje" que logran indexarse en Scopus y WoS*, ofreciendo una vía rápida para publicar sin revisiones exhaustivas y, en la mayoría de los casos, a cambio de pagos elevados. Estas revistas suelen cobrar entre *dos mil y cuatro mil dólares* por conceptos como edición, traducción y maquetado, transformando el proceso de publicación científica en una transacción comercial donde el criterio principal ya no es la calidad, sino la capacidad de pago de la persona responsable de la investigación.

Esta realidad pone en duda la fiabilidad de ciertos artículos, se cuestiona la eficacia de los filtros de indexación y se plantea un reto para las instituciones: garantizar que las publicaciones reconocidas realmente aporten conocimiento. Luego de examinar el ecosistema editorial legítimo, es necesario abordar su contraparte: las revistas depredadoras.



## ***Las revistas depredadoras: tipos, causas y consecuencias***

Revistas fraudulentas o de estafa.

En el otro grupo (Figura 1), identificado con el color violeta, se encuentran las denominadas revistas depredadoras o piratas, las cuales operan bajo modelos de publicación que comprometen la calidad y la integridad científica. Dentro de este grupo, un subgrupo particularmente preocupante es el de las revistas fraudulentas o de estafa, que emplean técnicas informáticas avanzadas para hackear y replicar el formato de revistas académicas legítimas con el propósito de engañar a las personas autoras y cobrarles por publicar en ellas (Jiménez-Yañez & Colmenares-Díaz, 2022).

De acuerdo con Hemmat et al. (2015), estas revistas fraudulentas presentan una serie de características distintivas: publican artículos sin un proceso de revisión por pares o con revisiones extremadamente laxas, sus páginas web suelen ser rudimentarias y carecen de información clara sobre su estructura y lineamientos editoriales, no tienen un enfoque científico definido ni especialización en un campo del conocimiento, el tiempo de aceptación y publicación de los artículos es inusualmente corto o, en algunos casos, inexistente, las tarifas de publicación no se especifican de manera pública, sino que se negocian en privado con los firmantes del paper, y utilizan tácticas agresivas de envío

masivo de correos electrónicos (spamming) para atraer a profesionales de la investigación desprevenidas.

Lukic et al. (2014) destacan que estos cibercriminales desarrollan estrategias dirigidas específicamente a las personas autoras que, presionados por las exigencias académicas, buscan publicar en un período de tiempo reducido. En muchos casos, estos investigadores, ya sea por desconocimiento, desesperación o falta de asesoramiento, caen en estas trampas y ven comprometida su reputación académica al asociarse con publicaciones de dudosa procedencia. Este fenómeno representa un desafío creciente en el ámbito de la comunicación científica, ya que la proliferación de revistas fraudulentas no solo afecta a investigadores individuales, sino que también distorsiona los sistemas de evaluación académica y debilita la credibilidad del conocimiento generado en distintas disciplinas.

### **Revistas sin respaldo institucional.**

El segundo subgrupo dentro de las revistas depredadoras está conformado por aquellas publicaciones que, si bien *no recurren al fraude informático*, presentan una serie de deficiencias estructurales y malas prácticas que comprometen la calidad del conocimiento científico. Estas revistas, en su mayoría, *carecen de respaldo institucional, acreditaciones, indexaciones y certificaciones reconocidas*. Además, su información de contacto no es verificable, lo que

dificulta la identificación de sus responsables. Otro aspecto distintivo es que no suelen estar indexadas en bases de datos de alta calidad, como Scielo, Scopus o Web of Science, y, en muchos casos, omiten completamente el proceso de revisión por pares, uno de los pilares fundamentales de la validación académica (Grzybowski et al., 2017).

El modelo de negocio de estas revistas se ha desarrollado a partir del auge del *Movimiento de Acceso Abierto*, aprovechando el sistema de pago por procesamiento de artículos (*Article Processing Charge* o APC, por sus siglas en inglés). A través de esta estrategia, obtienen ingresos cobrando a los firmantes de la investigación por la publicación de sus artículos sin garantizar estándares mínimos de calidad editorial ni revisión científica rigurosa (Abad-García, 2019).

Esta lógica de monetización, aunque responde a una creciente demanda por publicar, no puede entenderse al margen del contexto estructural del capitalismo académico. Slaughter y Leslie (1997) explican que el mercado ha reconfigurado la misión universitaria, desplazando el énfasis desde la enseñanza hacia la captación de fondos externos mediante investigación rentable, lo que ha erosionado el vínculo histórico entre academia y sociedad. Este cambio ha impulsado a las instituciones y al personal investigador a priorizar resultados cuantificables, como publicaciones e índices de impacto, por encima de la calidad o la utilidad social del conocimiento generado. En consecuencia, las

revistas depredadoras se insertan como una respuesta funcional a estas exigencias, ofreciendo visibilidad académica sin los filtros de calidad ni la ética editorial requeridos, lo que pone en riesgo la integridad del ecosistema científico.

Esta situación ha generado un debate en torno a la credibilidad de las revistas de acceso abierto en general, ya que la proliferación de prácticas cuestionables ha llevado a falsas generalizaciones que afectan incluso a editoriales legítimas. Un factor clave en esta percepción ha sido la influencia de la *lista de Beall*, un registro elaborado por el bibliotecólogo Jeffrey Beall, que recopilaba editoriales y revistas potencialmente depredadoras dentro del modelo de acceso abierto. A pesar de que esta lista fue criticada por su falta de transparencia y por depender del criterio de un solo individuo, su impacto en el campo de la bibliometría ha sido significativo, moldeando la forma en que se analizan y categorizan las publicaciones académicas en este ámbito (Krawczyk & Kulczycki, 2021).

## **Revistas "garaje" con apariencia de legitimidad**

El tercer grupo, denominados por el autor de este artículo como *revistas "garaje"*, representa una de las amenazas más perjudiciales y subestimadas para el sistema científico global. A diferencia de las revistas fraudulentas o depredadoras tradicionales, estas publicaciones logran acceder a indexaciones prestigiosas en bases de datos

reconocidas, lo que les otorga una apariencia de legitimidad. Sin embargo, en la práctica, operan bajo las mismas estrategias y deficiencias que las revistas depredadoras, poniendo en riesgo la integridad del conocimiento científico.

Las revistas "*garaje*" se caracterizan por la ausencia de un sistema riguroso de *revisión por pares*, aceptando artículos con evaluaciones superficiales o inexistentes. Además, publican una cantidad excesiva de números al año—frecuentemente más de cuatro—lo que indica un modelo de producción basado en la cantidad y no en la calidad. También tienden a aceptar un *número inusualmente alto de firmantes por artículo*, lo que puede sugerir la existencia de acuerdos cuestionables para favorecer la publicación de múltiples investigadores sin una evaluación estricta de su contribución real.

Otro rasgo distintivo de estas revistas es su modelo de financiamiento, basado en el cobro de múltiples tarifas a las personas autoras bajo distintos conceptos, como *edición, traducción o maquetación del manuscrito*. Estas tarifas, que en revistas científicas legítimas suelen ser transparentes y justificadas, en las revistas "garaje" se convierten en un mecanismo para lucrar sin ofrecer un verdadero servicio editorial de calidad.

El peligro de este tipo de revistas radica en que, al estar indexadas en bases de datos reconocidas, generan confusión

entre los investigadores, dificultando la identificación de publicaciones de alto impacto y rigor académico. Su presencia también debilita los sistemas de evaluación científica, ya que permiten que personas con escasa producción rigurosa acumulen publicaciones en revistas con apariencia de legitimidad, afectando el reconocimiento y la distribución de recursos dentro de la comunidad científica.

En este contexto, es fundamental desarrollar criterios más exigentes para la indexación en bases de datos prestigiosas, así como promover mecanismos de vigilancia que impidan que este tipo de publicaciones se perpetúen en el ecosistema académico global.

### **La distribución geográfica e impacto en citaciones de las revistas depredadoras.**

Con el paso del tiempo, las malas prácticas de las *revistas depredadoras* se han convertido en un problema creciente dentro de la comunidad científica. Demir (2018) llevó a cabo un estudio en 2017 basado en la lista de Beall, en el cual identificó que los tres países con el mayor número de publicaciones en este tipo de revistas eran *India, Nigeria y Turquía*. Estos países, según los estándares del Fondo Monetario Internacional (FMI), se encuentran en vías de desarrollo, lo que sugiere una correlación entre la precarización del sistema académico y la proliferación de publicaciones en revistas de baja calidad. Sin embargo, la

investigación también reveló que *países desarrollados*, como *Estados Unidos, Reino Unido, Japón y Corea del Sur*, figuraban dentro del *top 20* de naciones con más publicaciones en estas revistas, lo que evidencia que el problema no es exclusivo de economías emergentes.

Asimismo, Demir (2018) analizó la distribución geográfica de las revistas depredadoras a través de su *dirección IP*, encontrando que la mayoría estaban registradas en *India*, seguida por *Estados Unidos, Turquía y Reino Unido*. Esto indica que la presencia de estas revistas no solo responde a la demanda del personal investigador en busca de publicaciones rápidas, sino también a una oferta consolidada dentro del ecosistema editorial global, en el que algunos países han desarrollado verdaderas industrias de publicaciones depredadoras.

Otro aspecto de relevancia es el impacto que tienen estos estudios en el mundo científico. Frandsen (2017) analizó el nivel de citación de artículos publicados en revistas depredadoras en 2013 y encontró que fueron citados un total de *1,295 veces en los cuatro años siguientes*, una cifra considerablemente menor en comparación con la de una revista académica promedio. De manera similar, Björk et al. (2020) realizaron un estudio bibliométrico sobre *250 artículos publicados en 2014 en revistas depredadoras* y descubrieron que, cinco años después, el *promedio de*

*citaciones por artículo era de apenas 2.6, y que el 56% de los artículos no había recibido ninguna citación.*

Aunque el impacto de las revistas depredadoras es *relativo y no tan extendido*, su existencia sigue representando un desafío significativo para la *credibilidad y calidad del conocimiento científico*. La proliferación de estas publicaciones pone en riesgo la integridad académica, ya que facilita la difusión de estudios sin un adecuado proceso de *revisión por pares*, lo que puede generar información errónea o poco rigurosa en el ámbito académico. Además, el hecho de que algunos artículos publicados en estas revistas logren ser citados evidencia que, en ciertos casos, la *producción de baja calidad se infiltra en el circuito académico legítimo*, generando *ruido en la literatura científica* y dificultando la identificación de estudios con verdadero impacto. Esta situación compromete la fiabilidad de los sistemas de indexación y evaluación científica, debilitando la confianza en el proceso de publicación.

Estas prácticas dan lugar a consecuencias de mayor alcance, que trascienden el plano editorial y repercuten en la estructura del sistema científico global.

### **Consecuencias éticas y estructurales en el sistema académico.**



El impacto de las revistas depredadoras va más allá de la simple diseminación de información cuestionable. Un problema adicional es que *una parte del personal investigador recurre a estas publicaciones*, especialmente aquellas que cuentan con *alguna indexación de prestigio*, con el objetivo de *obtener promociones, incentivos económicos o beneficios en los ámbitos académicos y de estímulos públicos*. En muchos sistemas universitarios y gubernamentales, la cantidad de publicaciones sigue siendo un criterio central para la asignación de recursos, ascensos y reconocimiento profesional, lo que ha incentivado prácticas cuestionables.

*La falta de mecanismos de verificación efectivos por parte de los Estados e instituciones académicas* ha permitido que muchos académicos logren avanzar en sus carreras basándose en publicaciones de dudosa calidad. Esto no solo distorsiona los indicadores de producción científica, sino que también contribuye a la proliferación de estas revistas, ya que sigue existiendo una demanda sostenida de espacios donde publicar con rapidez y sin una evaluación rigurosa.

Por lo tanto, *es fundamental reforzar los sistemas de control y evaluación científica*, estableciendo criterios más estrictos para la indexación de revistas, promoviendo la educación ética en la publicación académica y fortaleciendo los mecanismos de supervisión por parte de agencias gubernamentales y universidades. Solo a través de una mayor regulación y una *cultura académica orientada a la calidad y*

*no solo a la cantidad de publicaciones, se podrá mitigar el impacto de estas prácticas y proteger la integridad de la investigación científica global.*

## **Conclusiones**

### *Causas y consecuencias del auge de las revistas de baja calidad*

El crecimiento de las revistas de baja calidad en el ámbito académico se debe a múltiples factores interrelacionados producto de causas individuales y colectivas. Entre las causas que destacan está la presión que ejercen las instituciones por publicar, la falta de una formación ética, la utilización de criterios cuantitativos como la medida substancial y principal de la productividad científica, y la lógica del capitalismo académico que la incentiva la acumulación de artículos por encima de la calidad del conocimiento generado. Estas circunstancias son una oportunidad que los responsables de las revistas depredadoras han sabido aprovechar, logrando que muchas de ellas incluso logren mantener una conexión temporal con las bases de datos de prestigio.

Esto ha traído una serie de consecuencias que derivan de este fenómeno. Entre ellas se encuentra la desconfianza en el sistema, la afectación de trayectorias académicas, el

rechazo de las instituciones, el debilitamiento de la ética profesional y el rating de las investigaciones.

### *Importancia del control de calidad en las bases de datos académicas*

Ante esta problemática, las bases de datos científicas de alto prestigio, como *Scopus*, *Web of Science (WoS)* y *Scielo*, cuentan con sistemas de control y seguimiento que garantizan la calidad de las revistas indexadas. Cada año, estas plataformas *examinan y excluyen revistas* que no cumplen con los requisitos mínimos para la publicación académica, como la rigurosidad en la revisión por pares, la transparencia en sus procesos editoriales y la calidad de sus artículos. Sin embargo, sus filtros de selección no son infalibles, afectando la credibilidad del conocimiento científico.

Uno de los riesgos para quienes optan por publicar en revistas depredadoras de dudosa procedencia es que estas van a ser *detectadas y expulsadas* de la base de datos. Esto genera una *pérdida significativa de la inversión realizada*, ya que los artículos publicados en estas revistas pierden su visibilidad e impacto académico. Además, en muchos casos, los artículos que fueron publicados en revistas eliminadas no pueden ser reubicados en nuevas publicaciones de mayor calidad.

## *Relevancia del ORCID en la credibilidad académica*

Toda revista académica sería exige como requisito el *ORCID (Open Researcher and Contributor ID)* de los firmantes de un manuscrito. Este identificador único permite validar la trayectoria de un profesional de la investigación y sus publicaciones, brindando transparencia y credibilidad a su trabajo.

El equipo editorial de revistas científicas legítimas revisan detalladamente el historial de publicaciones de los investigadores. Si en el *currículo científico* aparecen artículos en revistas depredadoras o de baja calidad, la *posibilidad de que un nuevo manuscrito sea aceptado disminuye considerablemente*. En este sentido, el prestigio y la reputación académica —los activos intangibles más valiosos de un científico— pueden verse *seriamente afectados* por decisiones de publicación erróneas.

## *Publicar: un proceso de aprendizaje y consolidación académica*

La publicación de artículos científicos no solo es un requisito profesional, sino también un *proceso de formación* que permite el desarrollo de *habilidades y competencias* fundamentales en la investigación. Desde el uso del lenguaje académico hasta la adaptación a la tipología y normativas de las revistas indexadas, cada experiencia editorial contribuye

al *fortalecimiento del perfil académico* de la persona responsable de la investigación.

Con el tiempo, quienes logran comprender y dominar el sistema de publicación científica incrementan sus *índices de éxito*, lo que les abre oportunidades para acceder a mejores revistas, aumentar su visibilidad y consolidarse en su campo de estudio.

### *Construcción de un perfil académico y oportunidades como revisor*

Un paso clave en la consolidación de un *perfil académico sólido* es recibir invitaciones para participar como *revisor externo en revistas de alto impacto*. Ser reconocido como revisor es un indicador de prestigio dentro de la comunidad científica y permite al investigador o investigadora expandir su red de contactos y fortalecer su credibilidad.

Sin embargo, un historial de publicaciones en revistas depredadoras o de baja calidad *puede cerrar esta puerta*. Hoy en día, una simple búsqueda en *Google o en ORCID* puede revelar el historial de un profesional de la investigación, y si su trayectoria muestra publicaciones en revistas cuestionables, será difícil que una revista de alto impacto confíe en su criterio para evaluar nuevos manuscritos.

## *Publicar como un proceso de desarrollo*

La publicación científica es un proceso riguroso que requiere *ética, estrategia y formación continua*. Optar por revistas de prestigio, cuidar la reputación académica y desarrollar un historial de publicaciones sólido no solo incrementa las oportunidades de éxito en la academia, sino que también evita riesgos innecesarios que pueden afectar la credibilidad de la persona responsable de la investigación a largo plazo.

Publicar en revistas científicas legítimas es un *proceso riguroso y desafiante*, pero también es una oportunidad invaluable para *desarrollar habilidades de investigación, fortalecer el pensamiento crítico y contribuir genuinamente al avance del conocimiento*. En este recorrido, el rechazo de un manuscrito no debe verse como un obstáculo insuperable, sino como *una parte natural del proceso académico* que permite mejorar, aprender y perfeccionar el trabajo.

La perseverancia, la integridad y el compromiso con la calidad y la ética científica son *pilares esenciales para construir una trayectoria académica sólida y respetada*. Solo mediante la dedicación constante y el respeto riguroso por los estándares de investigación es posible generar un impacto significativo en nuestras disciplinas y contribuir de manera genuina al avance del conocimiento global.

## *Desafío colectivo: hacia una ciencia ética y comprometida*

Finalmente, preservar la integridad del sistema científico no es responsabilidad exclusiva de los investigadores. Se requiere un compromiso colectivo que involucre a comités editoriales, revisores, instituciones académicas y entidades evaluadoras. Solo a través de una acción coordinada será posible enfrentar el auge de publicaciones de baja calidad y fortalecer el desarrollo científico para el bien común.

## **Referencias**

- Abad-García, M. (2019). El plagio y las revistas depredadoras como amenaza a la integridad científica. *Anales de Pediatría*, 90(1), 57.e1–57.e8. <https://doi.org/10.1016/j.anpedi.2018.11.003>
- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1), 377–386. [https://doi.org/10.1162/qss\\_a\\_00019](https://doi.org/10.1162/qss_a_00019)
- Bernal, C. (2016). *Metodología de la investigación. Administración, economía, humanidades y ciencias sociales* (4.ª ed.). Pearson Educación.
- Bertoglia, M., & Águila, A. (2018). Revistas depredadoras: Una nueva amenaza a las publicaciones científicas. *Revista Médica de Chile*, 146(2), 206–212.

<https://doi.org/10.4067/s0034-98872018000200206>

- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376. [https://doi.org/10.1162/qss\\_a\\_00018](https://doi.org/10.1162/qss_a_00018)
- Björk, B.-C., Kanto-Karvonen, S., & Harviainen, J. T. (2020). How frequently are articles in predatory open access journals cited. *Publications*, 8(2), 1–12. <https://doi.org/10.3390/publications8020017>
- Borgman, C. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. The MIT Press. <https://doi.org/10.7551/mitpress/7434.001.0001>
- Brunner, J. J., Labrana, J., Ganga, F., & Rodríguez-Ponce, E. (2019). Circulación y recepción de la teoría del "capitalismo académico" en América Latina. *Education Policy Analysis Archives*, 27, 1–79. <https://doi.org/10.14507/epaa.27.4368>
- Cortegiani, A., Ippolito, M., Ingoglia, G., Manca, A., Cugusi, L., Severin, A., Strinzel, M., Panzarella, V., Campisi, G., Manoj, L., Gregoretti, C., Einav, S., Moher, D., & Giarratano, A. (2020). Citations and metrics of journals discontinued from Scopus for publication concerns: The GhoS(t)copus Project. *F1000Research*,



<https://doi.org/10.12688/f1000research.23847.2>

Demir, S. (2018). Predatory journals: Who publishes in them and why? *Journal of Informetrics*, 12(4), 1296–1311.  
<https://doi.org/10.1016/j.joi.2018.10.008>

Departamento Administrativo de Ciencia, Tecnología e Innovación [Colciencias]. (2014). *Documento guía servicio permanente de indexación de revistas de ciencia, tecnología e innovación colombianas*.  
<https://minciencias.gov.co/sites/default/files/upload/paginas/M304PRO2G01-guiaserviciopermanente-indexacion.pdf>

Feather, J. (2003). *International encyclopedia of information and library science* (2.<sup>a</sup> ed.). Routledge.  
<https://doi.org/10.4324/9780203403303>

Frandsen, T. (2017). Are predatory journals undermining the credibility of science? A bibliometric analysis of citers. *Scientometrics*, 113(3), 1513–1528.  
<https://doi.org/10.1007/s11192-017-2520-x>

Grzybowski, A., Patryn, R., & Sak, J. (2017). Predatory journals and dishonesty in science. *Clinics in Dermatology*, 35(6), 607–610.  
<https://doi.org/10.1016/j.clindermatol.2017.07.003>

Hemmat, M., Wongwises, S., Asadi, A., & Akbari, M. (2015). Fake journals: Their features and some viable ways to distinguishing them. *Science and Engineering Ethics*,

21(4), 821–824. <https://doi.org/10.1007/s11948-014-9595-z>

Hostler, T. (2024). Open research reforms and the capitalist university: Areas of opposition and alignment. *Collabra: Psychology*, 10(1), 1–21. <https://doi.org/10.1525/collabra.121383>

Jessop, B. (2018). On academic capitalism. *Critical Policy Studies*, 12(1), 104–109. <https://doi.org/10.1080/19460171.2017.1403342>

Jiménez-Yañez, C., & Colmenares-Díaz, Z. (2022). ¿Qué se debe saber sobre las revistas depredadoras y piratas? *Culturales*, 10, 1–8. <https://doi.org/10.22234/recu.20221001.ed001>

Kendall, G. (2021). Beall's legacy in the battle against predatory publishers. *Learned Publishing*, 34(3), 379–388. <https://doi.org/10.1002/leap.1374>

Koerber, A., Starkey, J. C., Ardon-Dryer, K., Eko, L., & Kee, K. F. (2020). A qualitative content analysis of watchlists vs safelists: How do they address the issue of predatory publishing? *The Journal of Academic Librarianship*, 46(6), Article 102236. <https://doi.org/10.1016/j.acalib.2020.102236>

Krawczyk, F., & Kulczycki, E. (2021). How is open access accused of being predatory? The impact of Beall's lists of predatory journals on academic publishing. *The Journal of Academic Librarianship*, 47(2), 1–11. <https://doi.org/10.1016/j.acalib.2020.102271>

- Lukic, T., Blesic, I., Basarin, B., Ivanovic, B., Milosevic, D., & Sakulski, D. (2014). Predatory and fake scientific journals/publishers: A global outbreak with rising trend: A review. *Geographica Pannonica*, 18(3), 69–81. <https://doi.org/10.5937/GeoPan1403069L>
- Macháček, V., & Srholec, M. (2021). RETRACTED ARTICLE: Predatory publishing in Scopus: Evidence on cross-country differences. *Scientometrics*, 126(3), 1897–1921. <https://doi.org/10.1007/s11192-020-03852-4>
- Macháček, V., & Srholec, M. (2022). Retraction note to: Predatory publishing in Scopus: Evidence on cross-country differences. *Scientometrics*, 127(3), 1667. <https://doi.org/10.1007/s11192-021-04149-w>
- Marina, T., & Sterligov, I. (2021). Prevalence of potentially predatory publishing in Scopus on the country level. *Scientometrics*, 126(6), 5019–5077. <https://doi.org/10.1007/s11192-021-03899-x>
- Mertkan, S., Onurkan Aliusta, G., & Suphi, N. (2021). Knowledge production on predatory publishing: A systematic review. *Learned Publishing*, 34(3), 407–413. <https://doi.org/10.1002/leap.1380>
- Mills, D., & Inouye, K. (2021). Problematizing 'predatory publishing': A systematic review of factors shaping publishing motives, decisions, and experiences. *Learned Publishing*, 34(2), 89–104. <https://doi.org/10.1002/leap.1325>

- Saura, G., & Caballero, K. (2021). Capitalismo académico digital. *Revista Española de Educación Comparada*, 37, 192–210.  
<https://doi.org/10.5944/reec.37.2021.27797>
- Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Johns Hopkins University Press.
- Slaughter, S., & Leslie, L. (2001). Expanding and elaborating the concept of academic capitalism. *Organization*, 8(2), 154–161.  
<https://doi.org/10.1177/1350508401082003>
- Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy: Markets, state, and higher education*. Johns Hopkins University Press.  
<https://doi.org/10.56021/9780801879494>

# Understanding as a Requisite for Institutional Trustworthiness

*El entendimiento como un requisito para la  
confiabilidad institucional*

**Felipe Morales Carbonell<sup>102</sup>**

*Want of science, that is, ignorance of causes,  
disposeth, or rather constraineth a man to rely  
on the advise, and authority of others. For all  
men whom the truth concerns, if they rely not on  
their own, must rely on the opinion of some  
other, whom they think wiser than themselves,  
and see not why he should deceive them.*

*Ignorance of the signification of words; which is,  
want of understanding, disposeth men to take on*

---

<sup>102</sup> Universidad de Chile. <https://orcid.org/0000-0001-5492-0759>

*trust, not only the truth they know not; but also  
the errors; and which is more, the non-sense of  
them they trust: For neither error, nor non-  
sense, can without a perfect understanding of  
words, be detected. [...]*

*Ignorance of remote causes disposeth men to  
attribute all events to the causes immediate, and  
instrumental: for these are all the causes they  
perceive.*

— Thomas Hobbes, *Leviathan* (XI, pp. 17-18, 20)

*Well planted is not uprooted,  
well kept is not lost.*

— Lao Tzu, *Tao Te Ching* (54, trans. Ursula K. Le  
Guin)

## **Abstract**

Here, I examine the question of under what conditions institutions can be said to be trustworthy from the perspective of the epistemology of understanding. I argue that trustworthy institutions must understand the problems they address, manifest their understanding, and make themselves understandable to those who interact with them. Ideally,

institutions should be ‘open’ in the sense that third parties can integrate with them. Institutions should be engineered with these goals.

**Keywords:** Institutions; trust; understanding; social epistemology; open institutions.

## **Resumen**

Aquí examino la pregunta de en qué condiciones son confiables las instituciones desde la perspectiva de la epistemología del entendimiento. Argumento que para que las instituciones sean confiables, es necesario que entiendan los problemas a los que responden, manifiesten ese entendimiento, y se hagan entendibles para quienes interactúen con ellos. Idealmente, las instituciones deben ser ‘abiertas’, en el sentido de que sujetos ajenos a ellas puedan integrarse a su actividad. Las instituciones deberían diseñarse en vistas a estos problemas.

**Palabras clave:** Instituciones; confiabilidad; entendimiento; epistemología social; instituciones abiertas.

## **Introduction**

What is required for someone to be able to trust an institution? The question is not idle, as we are currently facing what some have called a crisis of trust in institutions. I start from the diagnosis that the crisis stems from, among other

factors, a failure on the part of institutions to act in ways that make them trustworthy. While others may think that this failure is essential to the nature of institutions, I think it is possible that this failure is merely contingent. We only have the institutions we have built, and we are imperfect at building them.<sup>103</sup> We may still improve our capacity to build institutions; and yet, in order to exercise this capacity, we might have to disassemble the institutions we currently have, and perhaps even reach out for new materials.

My interest here is merely to focus on certain issues that arise in the epistemology of trust, that directly touch on the issue of the trustworthiness of institutions. In at least certain conditions, to be spelled out later, institutions cannot be properly trusted *because* they present themselves as trustworthy in inappropriate ways, that is, because they *self-advertise*. Why are those ways inappropriate, and what does this show about the conditions for institutional trustworthiness? I will argue that in those cases the issue is that this form of self-advertisement is in conflict with transparency, because it masks the real trustworthiness of the

---

<sup>103</sup> We may lack the resources to build them better, but there may also not even be ways to perfectly solve the problem of designing institutions. In this second sense, the design of institutions may turn out to be a *wicked* problem (Rittel & Webber, 1973). Wicked problems don't have *neat* solutions because they cannot be formulated in a determinate way, have no *definite* solutions and are consequential and unique. Because they might be non-soluble, wicked problems require that we raise awareness, accept their insolubility and adapt around them.



self-advertising institution. Still, as some authors have recently pointed out, transparency may be itself problematic. I will argue that transparency is necessary for trust, but only insofar as it concerns the kind of understanding that the institution embodies concerning the sets of problems that it exists to solve. To elaborate on this idea, I will sketch an account of how institutions can embody understanding and manifest it and an account of social problems. I will then argue that we need to develop what I will call *open institutions*.

## **Robustness**

In the background of our discussion, we will find the *robustness* of institutions as an underlying concern. By robustness I mean the sensitivity of institutions to external pressures that could unravel them. Robust institutions can sustain themselves against attacks, the lack of a suitable environment in which to operate, contingent failure of the mechanisms that constitute them (or more generally, the contingent inability of the occupiers or roles within their structure), and so on.

Most importantly, we may think that at least some kinds of robust institutions do not require that the agents that interact with them hold certain attitudes of support towards them. This has a limit, since institutions are subject-dependent in the sense that their existence constitutionally

depends on the actions and attitudes of agents in a broad sense, but the point is that once they are set up, institutions can often sustain themselves to different degrees of independence from subjective support. In fact, that institutions can be robust in this way is usually taken as one of their main features: because they are relatively independent from circumstance, people can *rely* on them in a broader range of cases.

Reliance does not entail trust, even if trust may entail reliance.<sup>104</sup> Consequently, robustness does not entail trustworthiness. But robustness and trustworthiness are nonetheless linked. Distrust in institutions can undermine the functioning of institutions. Such undermining constitutes a failure of robustness in the broad sense I just described. Since there are incentives to prevent the unraveling of institutions, there are incentives to increase their robustness.<sup>105</sup> The same incentives push towards improving the trustworthiness of institutions, which would prevent distrust. However, this pair

---

<sup>104</sup> Here I bypass discussions about the nature of reliance and trust. There are accounts where trust is a form of reliance (so that it entails it), and accounts where it is not (so that we can have trust without reliance).

<sup>105</sup> Incentives for whom? For those who sustain the institutions, including the institutions themselves. This last point requires that we treat institutions as agents of their own kind, either primitively or through the mediation of whatever lower-level agents they might involve for their functioning. While I see no objection to taking institutions as primitive agents, I don't think what I say in this paper requires a determinate position on this point.

of incentives push also for the promotion of institutions *appearing* as trustworthy (even when they are not), *and* for institutions to be too insensitive to distrust.<sup>106</sup> Neither the appearance of trustworthiness nor robustness are good on their own. The former requires that the appearance of trustworthiness is grounded on true trustworthiness (in which case we say that it is *well-founded*), and the latter requires a balance between sensitivity and insensitivity. Institutions that fail to balance these demands are not trustworthy.

## **Self-advertisement**

Given this structure of incentives, where there are incentives to appear trustworthy in general, there are also incentives to make institutions appear trustworthy in every possible specific way. The incentives to do so in one way or the other, or to do so in one way rather than others, are in competition because of intrinsic incompatibility of the incentivized actions or by limited resources. Given a set of resources and a set of values, not every possible way to appear trustworthy will be pursued. Normatively, we should expect well-founded ways to appear trustworthy to be preferred to those that are not well-founded, but resource concerns and

---

<sup>106</sup> If taken to the extreme, these pressures lead to institutions that appear as trustworthy even though they are completely insensitive to whether any agent interacting with them trusts them.

other non-epistemic reasons can tip the balance in favor of the latter. When it is assessed that the resources that should be spent in appearing trustworthy in a well-founded manner are much too costly in comparison with non-well-founded ways, non-well-founded ways are preferred. This can be taken as a provisional measure, in a hope to eventually make the appearance of trustworthiness well-founded, but it is often taken without such prospects, in full knowledge that institutions are not properly trustworthy and that this erodes their functioning because they are unable or unwilling to take the steps to become trustworthy.

The paradigmatically Machiavellian strategy to use lies when expedient can be at play in such cases. The argument is one of risk minimization: noticing one's weakness, one displays strength or misdirects those one interacts with in a way that *conceals* those weaknesses.<sup>107</sup> This can be justified

---

<sup>107</sup> Thus, in Machiavelli: "Since a ruler has to be able to act the beast, he should take on the traits of the fox and the lion; the lion can't defend itself against snares and the fox can't defend itself from wolves. So you have to play the fox to see the snares and the lion to scare off the wolves. A ruler who just plays the lion and forgets the fox doesn't know what he's doing. Hence a sensible leader cannot and must not keep his word if by doing so he puts himself at risk, and if the reasons that made him give his word in the first place are no longer valid. If all men were good, this would be bad advice, but since they are a sad lot and won't be keeping their promises to you, you hardly need to keep yours to them. [...] Those best at playing the fox have done better than the others. But you have to know how to disguise your slyness, how to pretend one thing and cover up another. People are so gullible and so caught up with immediate concerns that a con man will always find someone ready to be conned. [...] So, a leader doesn't have to

appealing to the importance of preventing those weaknesses to be exploited: thus, for example, the state justifies its concealment by appeals to matters of security. People put up with it because they agree that the state plays an important part in such matters, and because they prefer to avoid trouble.

Self-advertisement can, thus, be seen as being in conflict with transparency, even though it may simply be a way to make trustworthiness visible. If an institution works hard to make itself trustworthy, it will be important to it that it can be seen as such; and indeed, that it is seen as trustworthy may not be an immediate consequence of being trustworthy. The key factor here is the ability of the people who interact with institutions to grasp how trustworthy they are. Even there, institutions have work to do, as I will argue.

## **Transparency**

Is the problem with self-advertisement that it can conflict with transparency? From our discussion so far, one may be tempted to think so. In particular, the risk-based considerations mentioned above suggest that self-

---

possess all the virtuous qualities I've mentioned, but it's absolutely imperative that he seem to possess them. I'll go so far as to say this: if he had those qualities and observed them all the time, he'd be putting himself at risk. It's seeming to be virtuous that helps; as, for example, seeming to be compassionate, loyal, humane, honest and religious. And you can even be those things, so long as you're always mentally prepared to change as soon as your interests are threatened" (2014, pp. 69–70).

advertisement is a way for self-advertisers to minimize risk by being less transparent.

As a strategy, being transparent has a complicated link to risk. If an agent is transparent about their own position in an environment (including information about their internal properties, and how they make them interact with the environment in the ways that they do) where other agents are not able to access information about them other than by the agent being transparent, and where there is no guarantee that other agents that are primed to grasp that information are aligned to the agent's goals, their being transparent can put them at risk: a non-aligned agent can use that information to improve their position in that environment over that of the agent, or to undermine the position of the agent in the environment. Signaling that one is willing to be transparent informs other agents that one is not worried about sharing information; it may weaken one's position on purpose in order to signal that one is willing to cooperate. So it makes sense that we wish institutions to be transparent, insofar as we hope institutions to be willing to cooperate. By weakening their position by being made transparent, the institution yields power to the agents that interact with it. To the extent that it is valued that people and not institutions are the ones that have political power, it makes sense to require transparency from institutions (of course, transparency is not sufficient for institutions to yield power to people). I take this to be the rationale for transparency *mandates* for institutions.

What is made transparent matters. In institutional contexts, transparency of the workings of institutions is particularly important, because it offers information about who does what in the institutions, and allows interested parties to hold institutions and its members *accountable*. Accountability is important because, in the cases where it allows us to detect who or what component of an institution misfires, it allows us to correct the issue without undermining the totality of the institution. In principle, accountability helps with robustness. And in cases where we detect that there is systemic trouble with the institution, (ideally) we get information about how to devise new institutions that replace the systems that are malfunctioning.

In summary, institutional transparency has two functions: one is *preemptive*, to incentivize institutions to yield power, and another is *forensic*, to determine accountability once the institutions has enacted the power that it has.

There has been some discussion about the justification for institutional transparency mandates recently.<sup>108</sup> This is Onora O'Neill, in her 2002 BBC Reith Lectures:

---

<sup>108</sup> Rose (1995) gives a general argument against monitoring, which transparency mandates can be seen as an instance of: “[...] constant monitoring and checking can poison the atmosphere. These activities have the appearance of a slur on the character of the monitored party. Indeed,

Transparency can encourage people to be less honest so increasing deception and reducing reasons for trust: those who know that everything they say or write is to be made public may massage the truth. Public reports may underplay sensitive information; head teachers and employers may write blandly uninformative reports and references; evasive and uninformative statements may substitute for truth-telling. Demands for universal transparency are likely to encourage the evasions, hypocrisies and half-truths that we usually refer to as ‘political correctness’, but which might more forthrightly be called either ‘self- censorship’ or ‘deception.’ (p. 73)

This focuses on the negative effects of transparency mandates. Skipping over the reference to ‘political correctness’, which has in the meanwhile turned into a dog-whistle,<sup>109</sup> O’Neill makes the point that transparency mandates create incentives for people to create the mere appearance of transparency. This is an argument against transparency mandates only insofar as these secondary incentives (not to be transparent, but to appear *as if* transparent) are the inevitable result of implementing mandates of this kind. This, however, is *prima facie*

---

the monitored party, feeling distrusted, may actually behave worse” (p. 540).

<sup>109</sup> Cf. Rinner & Hieke (2024).



implausible (we should eventually check whether it is *ultima facie* implausible as well). A more charitable take on the argument is that it captures the thought that the cost of creating these secondary incentives is higher than the cost of not mandating transparency. This, however, is unclear.

Nguyen (2021, 2026) offers a more nuanced version of this criticism. Instead of a blanket argument against transparency mandates, Nguyen focuses on the kind of transparency that is required from experts by the public. He provides two arguments to the effect that transparency mandates necessarily distort the actions of those experts of whom transparency is mandated. The first focuses on the intrusiveness of transparency mandates:

*Argument from epistemic intrusiveness*

- 1) Public transparency demands the the experts justify their actions to non-experts.
- 2) Many of the actual reasons for expert action are not available to non-experts.
- 3) The demand for transparency creates pressure for experts to offer justifications for their actions in terms available for the non-expert.
- 4) This pressure distorts, to varying degrees, expert action and communication.

The key idea here is that there is a differential of availability of reasons for different groups of people. A reason is available to members of a group if, in virtue of being members of that group, it is possible for them to grasp it, in the sense of being able to understand it as a potential justification for an action. Nguyen's argument is based on the idea that, because experts use a specialized conceptual framework, their reasons are unavailable to non-experts. This is explained by a certain division of cognitive labor: since not all members of society play the same roles in the epistemic processes of society, experts acquire epistemic competences that non-experts don't. Indeed, in Nguyen's view, it would be too costly for non-experts to acquire these competences: a society that divides labor is more effective than one that doesn't. The way out is to bridge the gap, and this can happen either by raising the competences of non-experts (which is costly and undermines the division of epistemic labor) or by narrowing the range of competences that experts deploy; it is in the latter case that transparency distorts the action of experts. This can happen in the form of deception on the part of experts (in which case it becomes misleading to say that they remain transparent), by their limiting their actions to those that can find public justification, or by preferring to act by publicly available reasons (in which case they essentially give up on their expertise). These forms of distortion fall into a gradation of intrusiveness, with deception on the lower end, and guidance distortion on the higher end. In the cases of

guidance distortion that Nguyen is worried about, experts leave aside their own skilled perception and expert intuition in favor of simplified rules of thumb or so-called ‘best practices’, “which usually take the form of clear and mechanically applicable guidelines for action, which offer deterministic recommendations for action based on publicly assessed criteria” (Nguyen, 2021, p. 13).

The second argument focuses on the importance of standpoint on the manner in which experts act and justify their actions:

*Argument from intimate reasons*

- 1) Some reasons are only available to people in some community or group.
- 2) So many of the actual reasons for action for members of that community are not available to outsiders.
- 3) The demand for transparency placed on that community—and its associated institution—asks that community members justify their actions in terms available to outsiders.
- 4) That demand distorts the actions and justifications of those community members.

This argument is more general in scope, as it applies to any group that has intimate reasons to act—reasons that are only accessible from their specific standpoint. The structure of the argument is sufficiently similar to that of the argument from epistemic intrusiveness, so there is no need for further discussion, except to note that Nguyen’s use of the notion of intimate reason suggests that the problem with transparency is deeper than previously assumed, as it not only distorts the reasoning of the agents who are mandated to be transparent in their quality of experts (and thus, as members of a community), but also their intimate reasoning, in their quality of possessor of intimate reasons. Since members of oppressed groups are possessors of such reasons, this suggests that transparency can be a means to oppression.

## **Understanding and Trust**

The previous sections have set up the stage. Nguyen, in particular, makes some comments that suggest a somewhat different diagnosis of the trouble with institutional trustworthiness in terms of transparency:

“The problem is that expert action is often deeply incomprehensible to the rest of us. In expert realms, corruption and the genuine pursuit of subtle goodness are very hard for the non-expert to tell apart” (Nguyen, 2021, p. 14).

The key here, I take it, is the appeal to comprehension. Something that is incomprehensible is difficult, and perhaps impossible, to trust. So perhaps we should say that trustworthy institutions must be comprehensible. In what follows, I will argue in favor of this idea.

The thesis that trustworthiness requires understandability seems plausible when we apply it to the case of individual people: *in order to put our trust in someone, we need to understand them*. In this section, I will examine this individual-level claim, in order to provide some background to my discussion of an institutional-level analogue claim I will discuss in the next section.

Why think that individual trustworthiness requires understandability? To answer this, I will draw from Stephen Grimm's (2016) account of understanding of people. Understanding, according to this proposal (and in) can take several forms, depending on the object of understanding. To understand why something happened, we need to grasp an explanation; to understand how something works, we need to grasp the structure of its operation; to understand a phenomenon, we understand the conditions of its occurrence, its consequences, etc. This form of understanding is, in Grimm's terminology, a kind of grasp of structure (often, causal structure), and it allows for inference, prediction, and manipulation/intervention. If we understand something in this way, we reduce its capacity to surprise us, and we can thus

come to rely on it being or behaving in certain ways and not others. Understanding people can also take this form. To understand why someone holds an opinion, we direct our gaze to the structure of the reasons they may have for holding it; to understand why someone has pursued crime, we may inspect the structural conditions that could influence their behavior. Grimm argues that, while this form of understanding people is important, it is not sufficient. We also need to grasp how come the goals that people pursue *can* be taken as worthwhile; we need to grasp their desirability. Grimm illustrates the point with an example from Anscombe (1957):

Suppose [...] someone were to claim that his overarching goal in life was to collect saucers of mud. Suppose too that his other beliefs and desires all aligned with this idea, so that they displayed [a] sort of internal consistency or coherence [...] even though there would be some sense in which we could discern structure in the person's actions (relations of dependence between beliefs, desires, and actions, e.g.) and thus achieve a kind of intelligibility or understanding, in another way the person's actions—guided by the dedication to saucers of mud—would seem extremely foreign and unintelligible. It would be a surd that would simply stare back at us. (Grimm, 2016, p. 217)

The reason why this matters (I think) is that, unless we also have this kind of understanding, we may fail to be able to predict people's behavior sufficiently well; in particular, whether we take different attitudes (such as trust) will be correct or incorrect depending on whether we grasp how it is that what other people do is worth pursuing for them. Understanding someone in the sense of seeing what they take to be good enables us to check whether our values are aligned or not. Trust requires that we can perceive this alignment; and in order to perceive that, we must have this kind of understanding.<sup>110</sup>

One may object that we often need to place our trust in people whom we do not understand. It is often the case that a child trusts their parents and tutors, even though they have only a superficial understanding of them. So clearly, there are cases of trust without understanding which serve as easy counterexamples to the thesis. In truth, these are not really counterexamples to the thesis. We must remember that understanding comes in degrees. A child trusts their parents and tutors *because* they understand them to some degree; not in detail, and perhaps not even accurately, but they still understand.<sup>111</sup> What they understand (or think they

---

<sup>110</sup> I will return to the issue of value alignment later.

<sup>111</sup> Cf. Grimm (2017) on some ways in which we can attribute understanding to children, and Clément (2010) on the capacity for children to trust. For some theories of understanding, children are

understand) is that their parents and tutors mean well with their actions, even if they fail to understand the actions themselves and the ways in which they may be good for them. Their trust is based on a partial understanding. If they lacked even that, they would not (indeed, they *do* not) put their trust in their parents and tutors.

The observation that understanding comes in degrees suggests that trustworthiness may also be a matter of degree. The more understandable someone is, the more trust we can place in them, *ceteris paribus*. This explains why we trust people who are aligned to us more easily than those who are not: they are more easily understandable for us.<sup>112</sup>

---

attributable with understanding only in an ‘honorific’ sense (cf. Kvanvig 2009). With Elgin (2009), I take that as a mistake.

<sup>112</sup> This seems to assume that we can understand ourselves and use our understanding of ourselves as a measure of how understandable others are. This may not be the case; cf. Williamson’s (2001) thesis that epistemic states are not ‘transparent’, so we may fail to be in a position to know that we are in them. This applies to both beliefs and desires. While Williamson does not mention the case of understanding explicitly, it is worth noting that in Williamson’s knowledge-first approach, understanding is probably itself a kind of knowledge, and Williamson *does* argue that knowledge is not transparent explicitly. While I don’t think understanding is a kind of knowledge, I acknowledge the issue concerning the non-transparency of epistemic states. In this case, however, we can bypass it noticing that we can make comparisons between ourselves and others without having full knowledge of either (similarly to how we can tune two strings without knowing the absolute pitch of either).



## Institutions and Understanding

I want to suggest that a similar story can be told for the institutional case. The motto would be: *only if you can understand an institution, you can trust it*. However, I want to argue that this actually has three distinct dimensions from the perspective of the conditions in which an institution can be trustworthy:

- 1) *Trustworthy institutions need to adequately understand the problems that they solve,*
- 2) *Trustworthy institutions need to act in ways that manifest their understanding,*
- 3) *Trustworthy institutions act in ways that make themselves understandable.*

These requirements, together, are stronger than the mere understandability condition. In particular, condition (1) is significantly more strict: an institution could be understandable without having itself any form of understanding. I want to say that such institution should not count as trustworthy. Because it has understanding, the ways in which displays understanding are well-founded. In turn,

our trust in institutions that satisfy these conditions is well-founded.<sup>113</sup>

It is important to notice that these conditions are not meant as sufficient for trustworthiness, not even together. As I will discuss later, an institution could satisfy them and still be untrustworthy. For now, I want to unpack what these conditions mean.

### ***Institutional Understanding***

The first condition requires that institutions are capable of understanding. In the individual case, there is no question as to whether individuals have this capacity. In the case of institutions, this is not so clear. Before anything else, we need an account of what it means to say that an institution understands.

The problem is compounded by the fact that institutions are not *groups*. If they were, we could perhaps claim that institutional understanding is reducible to some kind of aggregate of individual understandings, or appeal to some of the more sophisticated accounts of group epistemic

---

<sup>113</sup> I think this more stringent kind of condition also applies to the trustworthiness of individuals, but I will leave this merely as a suggestion here.

states in the literature.<sup>114</sup> In this case, institutional understanding would depend on the state of individuals. However, in the case of institutions, individuals are merely contingent occupiers of roles. I want to leave it open that an institution could be said to understand even if no individual occupied any role in the institution, although this is not essential to what follows.<sup>115</sup> In any case, I take institutions to be a kind of artifact, namely structures that constrain, determine and enact behaviors. Institutions enable or constrain the behavior of other actors, and interact with them as well. To do so, they must have some capacity to perform actions, and to do that, they need to grasp the procedures that structure those actions. The kind of understanding that we can attribute to institutions consists of this.

---

<sup>114</sup> These accounts usually build on top of the basis of a theory of group belief, acceptance or commitment. For an overview, see Tollefsen (2015).

<sup>115</sup> We could appeal to a distinction between “institutional facts” and “institutions proper” and say that institutions proper involve individuals but institutional facts do not necessarily involve them, but this strikes me as arbitrary in terms of the issue at hand (note that this notion of institutional fact is different from Searle’s (1995), which makes institutional facts depend on institutions—here, the class of institutional facts encompasses institutions). Institutional facts in this sense could be trustworthy or untrustworthy all the same. This interacts with the discussion on the nature of trust: if trust requires that one could be betrayed (Baier 1986), perhaps some institutional facts cannot be said to be trustworthy (it would be a category mistake to say that one could be betrayed by certain institutional facts). I am not fully convinced that the possibility of betrayal is constitutive of trust, although I happily concede that this possibility is crucial to understand many cases of trust.

For example, an institution may validate a certain kind of document. To validate a document has certain requirements, takes certain steps, and has some effects, which an actor who validates a document needs to grasp. The institution must grasp this heterogeneous but connected set of things, and the ways in which it is structured. To that extent, it must possess understanding of that structure. This understanding is encoded in the structure of the institution and its behavior; for example, there may be explicit norms that regulate the institution that it follows in its normal operation.

As artifacts, institutions are often created purposefully. The kind of institutions we assess as trustworthy or untrustworthy deal with practical matters that affect people (who put their trust or not in them). In general, institutions of this type are created to solve perceived problems, and designed to do so. Accordingly, the understanding that they have is understanding of the problems that they are designed to solve or that they at least as a matter of fact interact with. An institution is fit for purpose when it adequately deals with these problems, and unfit otherwise. This fit may be a matter of the intrinsic properties of the institution, or of the environment in which it operates; in either case, fit is a matter of degree. In any case, it is better, other things being equal, that the institution deals with the problems through understanding them than not. If institutions can have an

adequate understanding of these problems, they should have it.

An institution that is misled or fails to understand the problems that it is meant to solve cannot be trustworthy because it is not appropriately connected to the ground truth of the problems that it is meant to address. It will often try to intervene into a problem in ways that do not address it; if so, it will be ineffective and thus unreliable. If, by chance, it manages to intervene into a problem in ways that do address it, its success is accidental and thus not reliable and consequently, untrustworthy.

The requirement that trustworthy institutions should adequately understand the problems that they address rules out as untrustworthy institutions that merely appear as trustworthy: merely appearing to solve problems is not a way to solve problems (unless the problem is, precisely, to find out ways to appear to be solving problems, which may be the reason for some institutions to exist—there is a reason the phrase ‘solutions searching for problems’ exists).

It also rules out as untrustworthy institutions that, when mandated to be transparent, fall back into deception: the mandate of transparency is a way to address the problem of accountability that institutions face, and being deceitful is not a way to solve that problem but to dodge it. This illustrates the point that institutions need to understand that they are

addressing problems as institutions, on top of the specific problems that they are meant to address. An institution that fails to grasp that it is beholden to stakeholders, who may make claims of it, fails to be trustworthy in this same way. Because some problems are wicked, institutions that deal with them cannot be dogmatic about their own operation; they must be flexible and make compromises. An institution that deals with wicked problems that fails to grasp this is not trustworthy.

### ***How Institutions Manifest Understanding***

It is not sufficient that institutions in fact adequately understand the problems they are meant to address. Even if they do, they may not display this understanding in their normal operation, and thus not appear as trustworthy or appear less trustworthy than they should. So there are reasons to think that trustworthy institutions should manifest the understanding they have.

Institutions can only manifest their understanding through their operation, just like someone's understanding is manifested through their actions. In the individual case, many of the operations of understanding are overt, but there is also a private or covert dimension to understanding, where it manifests in a series of private episodes which are the occurrence of mental actions (someone *ponders* what to do, weighting arguments, assessing outcomes, and so on—the

kind of thing that *Le Penseur* is doing).<sup>116</sup> This dimension of individual understanding is manifested to other only through traces of these processes (the production of outputs from which it can be inferred that these private episodes took place, including retrospective reports about them). To speak of an inner mental life of institutions is at best metaphorical, but the operation of institutions also has overt and covert dimensions. The covert operation of institutions involves work that sustains its operation, but which is not shared with the public—for example, all the work that goes on in the offices of an institution. Some of this work gets traces, in the form of public announcements, reports, communication with other actors, overt actions, the effects of their actions, and so on. It is in these traces that institutions get the opportunity to manifest their understanding.

The claim was that trustworthy institutions should leave such traces, in a way that manifests understanding of the problems that they are dealing with. Because trustworthy institutions have adequate understanding of these problems, their understanding is particularly manifested in their successes, and more particularly yet, on their consistent success over time. While repeated success is not *certain* evidence of understanding, it is evidence nonetheless (note that it has to be weighted against evidence that shows that these successes might be attributable to mere luck, although

---

<sup>116</sup> Cf. Ryle (1968).

it is controversial whether understanding is incompatible with luck in the way that other epistemic states seem to be).<sup>117</sup> This display of successes is not the same as the kind of self-advertising we discussed earlier as problematic, because it is well-founded and is not meant as a way to conceal untrustworthiness. It is important that successes can be assessed by adequate public criteria, since that addresses the problem that those who interact with the institution may not be able to tell well-founded advertising from non-well-founded advertising otherwise. Still, the uptake of what institutions manifest of their understanding remains a problem.

Even trustworthy institutions can fail, so it is not necessary that the record of successes of institutions is immaculate. However, for trustworthiness it matters *how* they fail. Because failures yield information about their causes, trustworthy institutions spend efforts in learning from failures. That is, trustworthy institutions do not *just* fail—they build mechanisms for dealing with failures.<sup>118</sup> Conversely, what is a sure sign of untrustworthiness is the concealment of failure, which is a failure to properly manifest one's degree of understanding.

---

<sup>117</sup> Cf. Khalifa (2013). Hannon (2021) gives an overview of recent discussions.

<sup>118</sup> See Elgin (2017, ch. 14) and Morales Carbonell (2021).



The mandate for institutions to manifest their understanding is a mandate for transparency, but it differs from the kinds of mandates that O'Neill and Nguyen criticize. It is not *yet* a mandate for institutions to justify themselves in terms that the actors that interact with them can grasp, and it does not require that institutions guide themselves by criteria that are not grounded on their own understanding of the problems that they address. Nguyen's arguments mainly concern mandates for individuals who belong to certain social categories or roles to be transparent with regards to the output of their actions and the reasons they have to act as they do. Here, we are concerned with institutions being transparent, not necessarily with individuals being transparent. The way in which institutions interact with stakeholders is part of the kinds of problems that they are meant to address. So overt institutional guidance by transparency mandates is simply part of the ways in which institutions can manifest their understanding of the problems that they are meant to solve.<sup>119</sup>

---

<sup>119</sup> Perhaps similar obligations apply to individuals who occupy roles in institutions, in which case Nguyen's worry returns. One way to address it is to distinguish between those individuals as individuals and as role-occupiers, and say that they are bound to different sets of mandates in each sense. However, this move could lead to situations where the same individual is bound to conflicting duties. I do not know how to resolve this tension, or even if it is solvable.

## ***How Institutions Make Themselves Understandable***

As I pointed out above, even if an institution possesses understanding of the problems it addresses and manifests it, the public may not be able to perceive that it does, or to understand them. So it seems that institutions that in principle are trustworthy may not be trusted depending on things that escape their direct control. Nonetheless, institutions should make themselves understandable by the actors that interact with them, or at the very least take steps to make themselves understandable.

The assumption so far in the literature we have discussed is that institutions are responsible only for the information that they provide to the actors that interact with them, not for the capacities for uptake that these actors have. In this regard, institutions are taken as mere sources of information, not as actors in a more fully fledged way. I think this is a mistake. Institutions have, in principle at least, the capacity to teach skills to the actors that interact with them. We usually take this for granted, as interacting with institutions usually is mediated through systems that tell us what to do and how in order to interact with the institutions (for example, if I need to get a certificate from a certain institution, I might expect to be able to receive instructions on how to get it). These devices can, of course, fail—when they do, we see that the relevant parts of the institutional structure are not adequate for the institutional function of interacting

with other actors. This is reason to correct the institution, not the behavior of the actors that try to interact with them.

When actors lack the skills to interact with the institution, the institution should broadcast this to all parties that might influence those actors behavior, and make an attempt to at least promote the development of these skills on the part of actors who might interact with it in the future. Suppose that it is understood that for actors to interact with an institution, they need to have at least some minimal reading comprehension or knowledge, and it is assessed that the actors who are presently interacting with the institution lack these skills. Then, the institution is responsible to ameliorate this lack of skills, either by direct assistance or by the promotion and development of broad educational programs, either by the institution itself or by other institutions that might help with the problem. That many institutions lack mechanisms to implement such tasks shows that they are not fully committed to the norm to make themselves understandable.

In general, there is a certain asymmetry between the understanding that institutions have of the actors that interact with them and the understanding that these actors have of the institution. As the anthropologist James C. Scott (1998) has forcefully argued, institutions are very concerned

with making people legible, in the sense of being representable in ways that are suitable for action:<sup>120</sup>

Legibility is a condition for manipulation. Any substantial state intervention in society—to vaccinate a population, produce goods, mobilize labor, tax people and their property, conduct literacy campaigns, conscript soldiers, enforce sanitation standards, catch criminals, start universal schooling—requires the invention of units that are visible. [...] Whatever the units being manipulated, they must be organized in a manner that permits them to be identified, observed, recorded, counted, aggregated, and monitored. [...] the greater the manipulation envisaged, the greater the legibility required to effect it. (p. 183)

In Scott's account, states increase the legibility of society by imposing simplified schemes that interpret it under rigid categories that fail to account for the ways in which the people in society make sense of their practices, but which serve the function of simplifying statecraft.<sup>121</sup> These maps of society are meant to make society more understandable for

---

<sup>120</sup> Cf. Scott (1995).

<sup>121</sup> It is worth noting that it is precisely this kind of system of legibility that could promote transparency schemes, by imposing public criteria of assessment for actors in society who may have particular standpoints that could be distorted by the application of such criteria. In this respect, Nguyen's critique of transparency is spot on.

the institution of the state. Because they are set up in such a way that the state is organized towards making society match these schemes, they help make institutions insensitive to different ways to conceptualize society, undermining their own ability to adequately understand the problems that they are meant to solve.

On the other hand, as I have observed before, institutions can avoid becoming understandable to some of the actors that might interact with them, because that would relinquish control to those actors. This may suggest that we should push for institutions themselves to become more legible. In fact, institutions often try to make themselves legible to at least part of their own structure. This, however, does not mean that this makes institutions more understandable to themselves: it might reinforce misleading ideas about what these institutions are for, and how they should conduct themselves. In Scott's account, legibility requires a centralized organization, and it is not a given that such organization would foster (more) understanding.<sup>122</sup> Longino (1994), for example, argues that in science, the diffusion of power through the embrace of a variety of epistemic practices promotes the goals of inquiry.

---

<sup>122</sup> On this point, it might be worth considering the literature on cognitive diversity, which employs the notion of epistemic landscape (see O'Connor 2023, ch. 5 for an overview). The presence of centralized systems of coordination may foster or not the exploration of epistemic landscapes.

## **Social problems and value alignment**

I said that institutions qua artifacts are meant/designed to solve problems or address them as a matter of fact. The kinds of problems that they solve matters for trustworthiness.

Take an institution whose purpose is to enact violence on minorities, devised in response to a perceived problem with those minorities. Can such institution be trustworthy? I want to say no. The reason why it isn't should be, straightforwardly, that the problem that it addresses is of the wrong kind. Another reason to reject this institution as untrustworthy is that such institution would be morally reprehensible. These answers put us in the thorny realm of ethical discussion.

The underlying topic of both answers is, in fact, the same. To say that an institution solves the wrong problem is to take a stance of what counts as a problem worth addressing, and this is not a merely epistemic concern. In the case of institutions that address human problems, deciding what problems are worth addressing has ethical weight. Since there is bound to be disagreement, the problem becomes a matter of alignment and coordination of values.

In terms of alignment, it is implausible to think that any actor can treat as trustworthy an institution that it is not

aligned to their values. We can rely on institutions that oppose us, in the sense that we can predict their behavior and adjust our behavior accordingly. But we cannot trust them. In fact, these institutions can satisfy all the understanding-based conditions that we imposed on trustworthy institutions and still not count as trustworthy for us. I already said that these conditions are not meant to be sufficient for trustworthiness—rather, they are necessary conditions.

Still, we might want to push this further and turn these conditions into the characterization of a set of sufficient conditions for trustworthiness. One way to do it is to fix a class of problems as objectively in need to be addressed, and then say that trustworthy institutions need to address some of those problems through an adequate understanding of them. What we would have done through this maneuver is to push the ethical issue into the definition of the class of problems to address.

For this purpose we could draw from the literature on ‘social problems’.<sup>123</sup> The idea is roughly this: there is a class of problems that societies (in general, or each particularly) face that *need* to be addressed. This can be spelled out in several ways. Earlier conceptualizations of social problems did so in terms of notions such as those of deviance from normal

---

<sup>123</sup> See Kitsuse & Spector (1973), Schneider (1985) and Spector & Kitsuse (2001).

functioning, dysfunction, structural strain or harmful conditions. This made it seem as if social problems could be defined once and for all—and indeed, if this could be done it would ground the need for certain institutions (an arrangement of those institutions would yield an account of the necessary structure of society and political organization). Institutions that addressed such social problems would have an advantage in terms of being potentially trustworthy, since the importance of the problems that they address should be in principle understandable for everybody. More recent approaches, under the banner of constructivism, have emphasized the subjective nature of problems, moving onto conceptualizations of social problems that characterize them in terms of claims that members of society make in different putative conditions. This relativizes the perception of social problems, and raises the issue of the dynamics of the claims by different parties that point towards social problems. The issue is too vast to fully address here, but it is worth keeping in mind.

## **Conclusion**

In this paper, I have argued that trustworthy institutions should satisfy certain conditions that involve understanding: trustworthy institutions should have adequate understanding of certain problems, manifest this understanding, and make themselves understandable (both in the sense of making themselves visible, and in the sense of



supporting the capacities of actors who interact with them in order for them to understand them). There may be many different kinds of institutions that satisfy these conditions. Whether there are actual institutions that meet these criteria, or whether we can turn actual institutions that fail to meet them into institutions that meet them are open questions. When it comes to social structures, we have to take what I would call an *engineering* approach: we need to set up goals and design methods to reach them. Society faces many challenges, and institutions are tools that we can use to face them. To serve this purpose, institutions need to be flexible, diverse, revisable, and open. For that purpose, institutions need to be guided and foster by understanding, even if that yields their power and efficacy. In the trade-off between robustness and openness, the latter should be preferred. As Lao Tzu reminds us, “the hard and great go under, the soft and weak stay up.”

## References

- Anscombe, E. (1957). *Intention*. Harvard University Press.
- Baier, A. (1986). Trust and antitrust. *Ethics*, 96(2), 231–260.
- Clément, F. (2010). To trust or not to trust? Children's social epistemology. *Review of Philosophy and Psychology*, 1, 531–549. <https://doi.org/10.1007/s13164-010-0022-3>

- Elgin, C. (2009). Is understanding factive? In D. Pritchard, A. Miller, & A. Haddock (Eds.), *Epistemic value* (pp. 322–330). Oxford University Press.
- Elgin, C. (2017). *True enough*. The MIT Press.
- Grimm, S. (2016). How understanding people differs from understanding the natural world. *Philosophical Issues*, 26, 209–225. <https://doi.org/10.1111/phis.12068>
- Grimm, S. (2017). Understanding and transparency. In S. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 212–229). Routledge.
- Hannon, M. (2021). Recent work in epistemology of understanding. *American Philosophical Quarterly*, 58(3), 269–290. <https://doi.org/10.2307/48616060>
- Khalifa, K. (2013). Understanding, grasping, and luck. *Episteme*, 10(1), 1–17. <https://doi.org/10.1017/epi.2013.6>
- Kitsuse, J. I., & Spector, M. (1973). Toward a sociology of social problems. *Social Problems*, 20(4), 407–419.
- Kvanvig, J. (2009). *The value of knowledge and the pursuit of understanding*. Cambridge University Press.
- Longino, H. (1994). In search of feminist epistemology. *The Monist*, 77(4), 472–485.
- Machiavelli, N. (2014). *The prince*. Penguin Classics.

- Morales Carbonell, F. (2021). Mistakes as revealing and as manifestations of competence. *Synthese*, 198, 3289–3308. <https://doi.org/10.1007/s11229-019-02281-y>
- Nguyen, C. T. (2021). Transparency is surveillance. *Philosophy and Phenomenological Research*, 105(2), 331–361. <https://doi.org/10.1111/phpr.12823>
- Nguyen, C. T. (2026). *The score: How to stop playing somebody else's game*. Penguin.
- O'Connor, C. (2023). *Modeling scientific communities*. Cambridge University Press.
- O'Neill, O. (2002). *A question of trust: The BBC Reith Lectures 2002*. Cambridge University Press.
- Rinner, S., & Hieke, A. (2024). False friends in political dogwhistles. *Inquiry*, 1–15. <https://doi.org/10.1080/0020174X.2024.2368581>
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4, 155–169. <https://doi.org/10.1007/BF01405730>
- Rose, C. M. (1995). Trust in the mirror of betrayal. *Boston University Law Review*, 75, 531–558.
- Ryle, G. (2009). The thinking of thoughts: What is Le Penseur doing? In J. Tanney (Ed.), *Collected papers* (Vol. 2, pp. 494–510). Routledge. (Original work published 1968)
- Schneider, J. W. (1985). Social problems theory: The constructionist view. *Annual Review of Sociology*, 11, 209–229.

- Scott, J. C. (1995). State simplification. *Journal of Political Philosophy*, 3, 191–234.
- Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.
- Searle, J. (1995). *The construction of social reality*. Free Press.
- Spector, M., & Kitsuse, J. (2001). *Constructing social problems*. Transaction Publishers.
- Tollefsen, D. (2015). *Groups as agents*. Polity.
- Williamson, T. (2001). *Knowledge and its limits*. Oxford University Press.

# UHEdiciones



Instituto para el Desarrollo  
de la Cultura y de la  
Sociedad - IDECS



**CIEPE**

*Ética Política, Economía*



*Info*  
**Kratos**