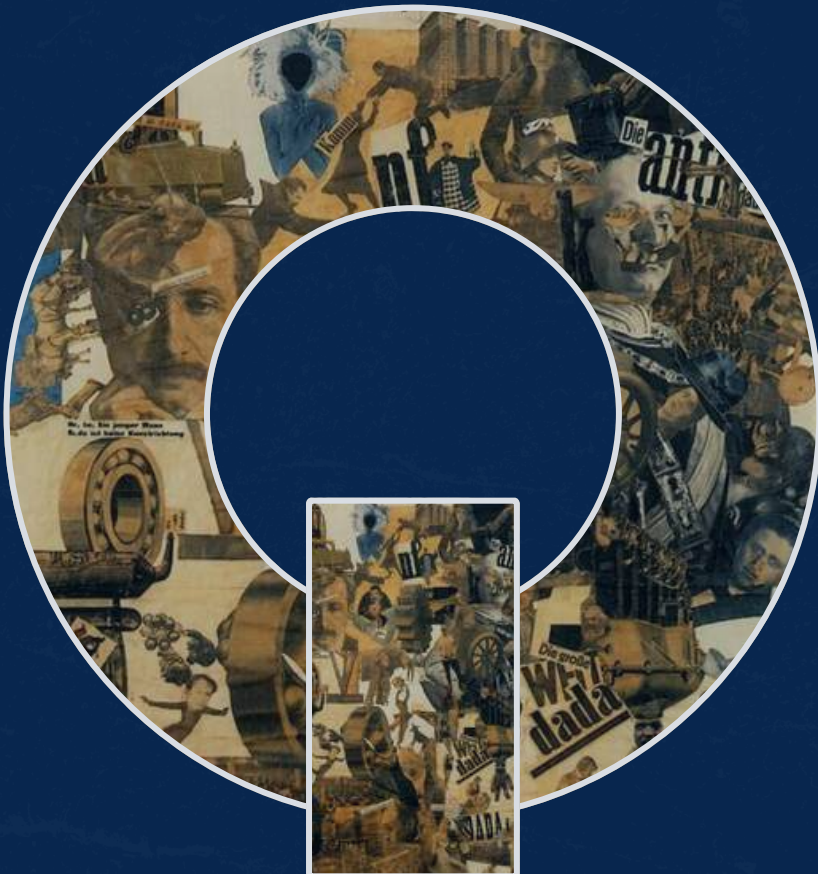


COLLOQUIA

"Revista de pensamiento y cultura"
"Academic journal of culture and thought"



Colloquia, Revista de Pensamiento y Cultura

Vol. 12

Enero – diciembre 2025

ISSN: 1390-8731

Periodicidad Anual

Indexado en:

Latindex: <http://www.latindex.org>

Advanced Science Index: <https://journal-index.org/index.php/asi/article/view/9560#>

European Reference Index for the Humanities and Social Sciences: <https://kanalregister.hkdir.no/publiseringskanaler/erihplus/>

Matriz de Información para el Análisis de Revistas: <https://miar.ub.edu/>

International Services for Impact Factor and Indexing (ISIFT): <http://journalimpactfactor.in/index.php>

International Institute of Organized Research (I2OR): <http://www.i2or.com/home.html>

Dialnet: <https://dialnet.unirioja.es>

Dirección

Andrea Puente León

Coordinación Editorial

Santiago Ullauri Betancourt

Comité Editorial

Andrea Puente León, Universidad Hemisferios. Ecuador

Ana Isabel Moscoso Freile, Universidad Hemisferios. Ecuador

Gabriela Baquerizo, Universidad de los Andes. Chile

Michelle Datiles, Catholic University of America. USA

Comité Científico

Mariano Fazio, Pontificia Universidad de la Santa Cruz. Italia

Jaime Nubiola, Universidad de Navarra. España

Enrique Banús, Universidad de Piura. Perú

Javier de Navascués, Universidad de Navarra. España

Ciro Parra, Universidad de la Sabana. Colombia

Diego Alejandro Jaramillo, Universidad Hemisferios. Ecuador

Paulina Dueñas, Universidad Hemisferios. Ecuador

Diagramación y diseño

Santiago Ullauri Betancourt

Juan Esteban Echeverry Murillo

Instituto para el Desarrollo de la Cultura y de la Sociedad

Universidad Hemisferios

UHEdici3nes

Paseo de La Universidad Nro. 300 y Juan Díaz (Urbanización Iñaquito Alto), Quito, Pichincha, Ecuador

Tel: (593)-2-4014-230

<https://colloquia.uhemisferios.edu.ec/index.php/colloquia/index>

colloquia@universidad.uhemisferios.edu.ec

Índice

Editors' note

Pablo Ruiz Osuna
Isaac Carbajal

Dossier: Artificial Intelligence: Convergences, Challenges, and Transformations in the Contemporary World

1.	Cognitive offloading and the reshaping of human thought: The subtle influence of Artificial Intelligence	Vincent J. Hooper	01–14
2.	Artificial Intelligence and the erosion of community ties: A philosophical reading based on the thought of Byung-Chul Han	Boris Osvaldo Saavedra Pérez	15–38
3.	Artificial intelligence literacy model as a complement to information literacy	Alejandro Villegas-Muro	39–70
4.	Generative AI for the preparation of reports in forensic psychology: Ethical, technical, and legal challenges	Guillermo Ramírez Zavala	71–105
5.	Generative AI and advertising campaigns: Social, ethical, and regulatory challenges in the algorithmic transformation of corporate communication	Fernando Checa García Víctor Núñez Fernández	106–134
6.	School leadership, Artificial Intelligence, and the ethical imperative in the STEM approach	Johanna Solange Godoy Cerda	135–157
7.	A Copernican shift for AI: From informational ethics to ontogenic responsibility	Raúl López Ramírez	158–183
8.	Generative AI in linear programming teaching: Errors, learning, and collective construction	Claudia E. Carignano Mariana Guardiola Evelín M. Rabbia Elena Rojas Heredia	184–200
9.	Ethical and anthropological status of artificial intelligence	Eduardo Gutiérrez Gutiérrez	201–246
10.	Beyond the result. Didactic reflection to identify and mitigate algorithmic biases	Arturo Muñoz Colunga	247–261
11.	The university and artificial intelligence: Critical tensions from transhumanism and posthumanism	Lorena Litai Ramos Luna	262–287

12.	Perceptual thickness: An approach to algorithmic landscape from art theory	Constanza Rivano Delzo	288–308
13.	Artificial Intelligence and its impact on cognitive work from a humanistic perspective	María Aránzazu Serantes	309–319
14.	The global system of Generative Artificial Intelligence. Work, power, and challenges in the second wave of information	Patricio Feldman Martín Gendler	320–353
15.	Religion and New Technologies: Technology with soul: contributions and limitations from a religious perspective	Alejandro Cerda Sanhueza	354–375
16.	Thinking despite the song of the cicadas	Héctor Inzaurrealde	376–388
17.	Artificial intelligence and computational creativity: The role of the artist in the posthumanist era	Iván Vanioff	389–415

Articles

18.	Beyond the market and the minimal state: The richness of liberal political philosophy	Guillermo Gastón Colella	416–441
19.	Deontic powers and social epistemology: An approach to testimony based on John Searle's philosophy of society	Felipe Álvarez	442–469
20.	Philosophical anthropology of memory: Background from Lluís Duch to the cultural formation of the anthropolos	Carolina Lagos Oróstica	470–485
21.	The concept of beauty in Thomas Aquinas: Between the metaphysical and the divine	Javier Ignacio Fattah Jeldres Cecilia Soledad Pérez Mora	486–507
22.	Impartiality in the ethics of Adam Smith and Immanuel Kant as the foundation of public reason and political liberalism	Nicolás Matías Fuentes Valdebenito	508–534
23.	Artificial Intelligence in the formulation and customization of public policies: Between efficiency, equity, and democratic risk	Dayanara Noelly Solano Carpio Boris Raúl Ochoa Ordóñez	535–557
24.	Emotions and politics: A theoretical and methodological strategy for analyzing the emotionalization of political discourse in the post-democratic context	Ekhiñe Graell Larreta Edith Flores Pérez	558–582
25.	Religious diversity in the Americas: A brief overview of hybridization, resistance, and revitalization	Tatiana Cabrera Mogrovejo Oswaldo Suin Guaraca	583–605
26.	Nineteenth-century anarchism confronting religion	Boris Briones Soto	606–629

EDITORS' NOTE

Nota de los editores

En un mundo atravesado por aceleradas transformaciones tecnológicas, la inteligencia artificial (IA) se ha convertido en uno de los fenómenos más decisivos y disruptivos del presente. Pensar la IA ya no es un lujo teórico ni un asunto exclusivo de especialistas técnicos: es una urgencia humana, social y política. La velocidad con la que los sistemas inteligentes reconfiguran prácticas cotidianas, instituciones y formas de relación exige una reflexión atenta, capaz de situar esta tecnología en su complejidad y de pensar sus múltiples efectos. Ante un horizonte que cambia sin cesar, la tarea intelectual se vuelve indispensable para comprender qué clase de futuro estamos comenzando a habitar.

Reflexionar sobre la IA es, por tanto, un ejercicio profundamente humano: interroga nuestros modos de actuar, de conocer y de construir sentido. En una época marcada por promesas de automatización y eficiencia, la mirada crítica se vuelve un acto de responsabilidad, un modo de no abdicar del pensamiento en medio del vértigo digital. La nostalgia por un pasado idealizado, aunque comprensible, resulta cada vez más insuficiente para afrontar los desafíos contemporáneos. La IA no solo transforma sistemas productivos o procesos institucionales; también reconfigura nuestra percepción del tiempo, del trabajo, de la subjetividad y de la convivencia social.

Frente a este panorama, el dossier **“Artificial Intelligence: Convergences, Challenges, and Transformations in the Contemporary World”** se propone reunir aportes que permitan pensar la IA desde múltiples aristas, reconociendo la necesidad de un diálogo interdisciplinario que abarque lo humano, lo social, lo político, lo filosófico, lo jurídico y lo económico. Al mirar de frente al futuro, sin temor y sin evasiones, aspiramos

a construir un espacio crítico y plural que interrogue las tensiones, riesgos y posibilidades abiertas por la IA, con especial atención a los derechos humanos, la ética centrada en la persona y los marcos conceptuales que nos permiten comprender sus transformaciones en la sociedad contemporánea.

La IA no puede seguir siendo comprendida como un terreno exclusivo de ingenieros, programadores o corporaciones tecnológicas. Aunque su desarrollo técnico sea indispensable, las decisiones que se toman en laboratorios y centros de innovación —muchas veces situados en Silicon Valley u otros polos tecnológicos, repercuten directamente en la manera en que vivimos, trabajamos, nos relacionamos y pensamos. Por ello, la IA debe abrirse a un examen que incorpore perspectivas filosóficas, sociales, antropológicas, jurídicas y políticas, pues su impacto desborda con creces el ámbito meramente computacional. Pensar la IA desde múltiples miradas no es una opción metodológica: es una necesidad cultural.

Comprender las implicancias de la IA exige, entonces, reconocer que cada línea de código arrastra concepciones del mundo, valores y modelos de sociedad. Los sistemas algorítmicos que parecen neutros en realidad portan decisiones que afectan derechos, hábitos, afectos y formas de vida. Aquello que se diseña en Silicon Valley se filtra en nuestras rutinas más cotidianas, desde la manera en que consumimos información hasta las formas de interacción y vigilancia. De ahí que la reflexión interdisciplinaria sea imprescindible: solo una mirada plural y rigurosa puede desactivar ingenuidades, revelar tensiones y orientar la discusión hacia horizontes más humanos y más justos.

La pertinencia de este dossier radica en la necesidad urgente de generar espacios académicos capaces de interpretar críticamente la expansión de la IA y sus consecuencias en la vida contemporánea. En una época en que los desarrollos tecnológicos se despliegan con rapidez y alcanzan dimensiones globales, las sociedades requieren marcos conceptuales que permitan comprender, evaluar y orientar estos procesos. La producción científica no puede permanecer al margen: tiene la responsabilidad de iluminar zonas de

tensión, identificar riesgos emergentes y proponer caminos de reflexión que acompañen y cuestionen el avance tecnológico.

Asimismo, este dossier resulta pertinente porque busca reunir investigaciones que dialoguen entre diferentes disciplinas, rompiendo con la fragmentación habitual entre ciencias humanas, sociales y tecnológicas. La IA se sitúa en una encrucijada donde convergen problemas éticos, jurídicos, económicos y culturales, por lo que exige un enfoque plural capaz de captar su complejidad. Al promover esta convergencia interdisciplinaria, el dossier aspira a enriquecer el debate contemporáneo y a ofrecer herramientas teóricas y metodológicas que respondan a desafíos colectivos cada vez más apremiantes.

Finalmente, la relevancia de este dossier se sustenta en su vocación de intervenir en la discusión pública sobre los horizontes de futuro que se están configurando. Las decisiones que se tomen hoy respecto al desarrollo y la regulación de la IA influirán de manera determinante en la protección de derechos humanos, la estructura de nuestras instituciones democráticas y la configuración de nuestras formas de vida. En este sentido, el dossier no solo pretende analizar las transformaciones en curso, sino también contribuir a una conversación crítica que permita imaginar futuros más inclusivos, justos y humanos.

La creciente *siliconización* del mundo, como advierte Eric Sadin (2018), amenaza con reducirnos a usuarios pasivos de sistemas que operan según lógicas opacas y decisiones tecnológicas que no hemos contribuido a imaginar. Ante este escenario, los aportes académicos sobre IA se vuelven un acto de resistencia intelectual: permiten comprender las fuerzas que moldean nuestro presente y, con ello, abren la posibilidad de no limitarnos a consumir tecnología, sino de participar activamente en su significado y en sus consecuencias. Pensar la IA es también pensar los modos en que se configuran nuestras formas de vida, nuestras instituciones y nuestras subjetividades en un tiempo profundamente mediado por lo digital.

Cada disciplina está llamada a reflexionar sobre su futuro antes de que otros lo hagan por ella. Si dejamos la imaginación del porvenir exclusivamente en manos de corporaciones tecnológicas o actores económicos, renunciamos a la posibilidad de orientar colectivamente el rumbo de nuestras sociedades. Las humanidades, las ciencias sociales, el derecho, la economía, la filosofía o la política tienen la responsabilidad de interrogar cómo la IA reconfigura sus categorías fundamentales y qué desafíos abre para sus propios campos. Este ejercicio reflexivo no es una simple proyección futurista, sino una toma de posición frente a un presente en continua mutación que exige pensamiento crítico y creatividad conceptual.

En este sentido, resulta necesaria una visión humanista, crítica, a la vez que esperanzadora e integradora, que nos prepare para el mundo venidero (el conocido ya como la 4 revolución industrial o la era de las máquinas inteligentes). Ante *la ola que viene* parafraseando a Suleyman (2023), es inexcusable que las humanidades no eximan su necesaria vocación de mantener una nueva narrativa que permita la convivencia con lo artificial, toda vez que se mantiene a *lo humano* en el centro.

Pensar juntos el futuro —desde la pluralidad de saberes— es indispensable para comprender que la tecnología nunca es neutra, tal como lo señaló Heidegger (1994). Toda innovación incorpora decisiones, valores e interpretaciones del mundo que orientan su impacto en la vida humana. Por ello, la construcción colectiva de un futuro posible implica examinar las oportunidades y riesgos que la IA despliega, reconociendo que su desarrollo debe estar anclado en debates éticos, políticos y culturales. Solo así podremos imaginar horizontes donde la tecnología esté al servicio de la transformación social, la justicia y la dignidad humana, y no de una automatización del sentido que nos desplace de nuestra propia capacidad de decidir.

El progreso tecnológico suscita siempre desafíos, y la IA, por su rápida aceleración e implementación a nivel mundial (se calcula que en 2025 ChatGPT tiene del orden de 800 millones de usuarios activos semanales (Singh, 2025)), plantea riesgos a todos los niveles (gobernanza, accesibilidad, derechos de las minorías, democratización de su implementación, etc.).

En este contexto de transformaciones vertiginosas, se vuelve urgente volver la mirada hacia nosotros mismos. Antes que la IA redefina por completo nuestras coordenadas de existencia, necesitamos comprender cómo estamos habitando este presente y qué tipo de humanidad queremos proyectar hacia el porvenir. Ver el futuro a los ojos implica reconocer su complejidad y, como sugiere Yuk Hui (2020), fragmentarlo: desarmarlo en sus múltiples capas para evitar las narrativas totalizantes que pretenden imponer un único camino tecnológico. Solo así podremos recuperar nuestra capacidad de agencia frente a un mundo que tiende a automatizar incluso nuestra percepción del tiempo y del sentido.

Pensar un futuro donde la IA será parte estructural del desarrollo humano exige imaginación, reflexión crítica y una profunda responsabilidad intelectual. La presencia creciente de sistemas inteligentes afectará nuestras formas de trabajo, nuestros vínculos, nuestras decisiones políticas y hasta nuestra comprensión de lo que significa ser persona. Este escenario no puede ser dejado al azar ni a los intereses de unos pocos. Imaginar y pensar son, por ello, actos de resistencia: modos de habitar activamente un futuro que ya está en gestación y del cual no podemos —ni debemos— permanecer al margen. El pensamiento no es decoración: es infraestructura del porvenir.

A medida que la ciencia ficción parece volverse cada vez más real, se vuelve indispensable llenar el futuro con nuestras ideas, nuestras preguntas y nuestras posibilidades. No para controlar lo inevitable, sino para acompañarlo con lucidez, diversidad y humanidad. La IA no es un destino escrito, sino un campo abierto que demanda la participación de todas las disciplinas y de todos los imaginarios. Este dossier se inscribe justamente en esa tarea: ofrecer un espacio donde el pensamiento pueda anticipar, cuestionar y reinventar el mundo que viene. Porque el futuro no se espera: se piensa, se construye y se habita.

La IA está destinada a llevarnos a un futuro prometedor. Cada vez nos damos cuenta con mayor acierto que estamos ante algo que supera (y con creces) la fenomenología de una mera herramienta o tecnología, acercándonos a unos pasos cada vez más acelerados a una nueva realidad: la

sociedad 5.0. Máquinas que crean, que piensan, que se relacionan con nosotros de un modo que produce vértigo, a la vez que nos obliga a repensar qué significa ser humano en la era digital.

Este Dossier representa un primer acercamiento sincero y multidisciplinar que ahonda en la necesidad temprana de no desoír el llamado que tiene todo académico de su época de preservar los intereses de la sociedad, toda vez que acompaña a los cambios sociales y tecnológicos del necesario e ineludible sentido que el presente y el futuro nunca deben desoír: preservar la dignidad humana.

Resumiendo someramente lo que hemos expresado, y como hemos advertido en más de una ocasión: “la IA lo cambiará casi todo”. Paulatinamente, todo sector (incluso aquellos que nunca creían que se verían afectados) acabará, irremediabilmente, conviviendo con la IA. No hay que temer al futuro, pero como advirtió Hans Jonas (1995), el principio de responsabilidad obliga siempre a mantener una mirada cauta y de eterna vigilancia, pues no se deben postergar las preguntas, los protocolos, las reflexiones críticas, a un futuro indeterminado. Esto es así, quizá, porque cuando hablamos de IA, y dadas sus vertiginosas e imparables aplicaciones, el futuro comienza hoy.

Pablo Ruiz Osuna & Isaac Carbajal
Editores Invitados

Referencias

- Heidegger, M. (1994). La pregunta por la técnica (E. Barjau, Trad.). En *Conferencias y artículos* (pp. 9–37). Serbal.
- Hui, Y. (2020). *Fragmentar el futuro*. Caja Negra.
- Jonas, H. (1995). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica*. Herder.
- Sadin, É. (2018). *La inteligencia artificial o el desafío del siglo: Anatomía de un antihumanismo radical* (J. Colomer, Trad.). Caja Negra.

Singh, S. (2025, 20 de noviembre). ChatGPT users stats (December 2025) – Growth & usage data. *DemandSage*. <https://www.demandsage.com/chatgpt-statistics/>

Suleyman, M., & Bhaskar, M. (2023). *La ola que viene: Tecnología, poder y el gran dilema del siglo XXI* (C. Fernández Morenas, Trad.). Debate.

DOSSIER:
ARTIFICIAL INTELLIGENCE:
CONVERGENCES, CHALLENGES,
AND TRANSFORMATIONS IN
THE CONTEMPORARY WORLD

Cognitive offloading and the reshaping of human thought: The subtle influence of Artificial Intelligence

La descarga cognitiva y la remodelación del pensamiento humano: La sutil influencia de la inteligencia artificial

Vincent J. Hooper

vincent.hooper@spjain.org

SP Jain School of Global Management

ORCID ID 0000-0001-6473-700X

Abstract: The integration of artificial intelligence (AI) into daily cognitive tasks has transformed traditional cognitive offloading—the delegation of mental processes to external tools—into a dynamic partnership with intelligent systems. This article examines the dual role of AI-driven cognitive offloading, exploring its potential to enhance efficiency and creativity while posing risks to memory consolidation, critical thinking, and intellectual autonomy. Grounded in extended mind theory and empirical studies like the Google Effect, the analysis reveals how AI's generative capabilities (e.g., ChatGPT, Midjourney) shift offloading from passive storage to delegated thinking, where users adopt AI outputs with minimal scrutiny, fostering automation bias. In educational contexts, AI tools risk undermining deep learning by reducing retrieval practice and encouraging superficial engagement, as illustrated by a case study of Hatt University, where AI-assisted essays distorted grading systems and eroded student critical analysis. Similarly, in healthcare and finance, overreliance on AI recommendations may compromise professional judgment. To mitigate these risks, the article proposes strategies such as metacognitive training, explainable AI design, and curriculum reforms prioritizing active engagement with AI outputs. Ethical and policy interventions are urged to address epistemic opacity, intellectual property, and the cultural redefinition of authorship. The study underscores the need

for balanced AI integration—one that harnesses its benefits while safeguarding human cognitive autonomy.

Keywords: Artificial intelligence; cognitive offloading; memory; critical thinking; creativity; automation bias; education; ethics.

Resumen: La integración de la inteligencia artificial (IA) en las tareas cognitivas diarias ha transformado la descarga cognitiva tradicional —la delegación de procesos mentales a herramientas externas— en una asociación dinámica con sistemas inteligentes. Este artículo examina el doble papel de la descarga cognitiva impulsada por IA, explorando su potencial para mejorar la eficiencia y la creatividad, al tiempo que plantea riesgos para la consolidación de la memoria, el pensamiento crítico y la autonomía intelectual. Basado en la teoría de la mente extendida y en estudios empíricos como el Efecto Google, el análisis revela cómo las capacidades generativas de la IA (por ejemplo, ChatGPT, Midjourney) desplazan la descarga cognitiva del almacenamiento pasivo hacia un pensamiento delegado, en el cual los usuarios adoptan las respuestas generadas por IA con una mínima revisión, fomentando el sesgo de automatización. En contextos educativos, las herramientas de IA corren el riesgo de socavar el aprendizaje profundo al reducir la práctica de recuperación y promover un compromiso superficial, como se ilustra en un estudio de caso de la Universidad Hatt, donde los ensayos asistidos por IA distorsionaron los sistemas de evaluación y erosionaron el análisis crítico de los estudiantes. De manera similar, en la atención sanitaria y las finanzas, la dependencia excesiva de las recomendaciones de IA puede comprometer el juicio profesional. Para mitigar estos riesgos, el artículo propone estrategias como la formación metacognitiva, el diseño de IA explicable y reformas curriculares que prioricen la interacción activa con las respuestas generadas por IA. Asimismo, se exhorta a implementar intervenciones éticas y de política pública para abordar la opacidad epistémica, la propiedad intelectual y la redefinición cultural de la autoría. El estudio subraya la necesidad de una integración equilibrada de la IA, que aproveche sus beneficios sin comprometer la autonomía cognitiva humana.

Palabras clave: Inteligencia artificial; descarga cognitiva; memoria; pensamiento crítico; creatividad; sesgo de automatización; educación; ética.

Introduction

In an era dominated by technological acceleration, artificial intelligence (AI) has shifted from being a mere auxiliary tool to an integral partner in human cognition. AI is embedded in daily activities such as writing, brainstorming, scheduling, decision-making, and even creative production. This article explores the phenomenon of cognitive offloading—the delegation of cognitive tasks to external agents or tools (Risko & Gilbert, 2016)—within the AI era. While cognitive offloading is not inherently negative, its pervasive use has profound implications for memory, critical thinking, and the very architecture of human thought.

This article advances a central claim: AI-driven cognitive offloading fundamentally restructures the architecture of human deliberation in three interconnected ways. First, it transforms attention allocation by reducing the cognitive effort directed toward information retrieval and synthesis. Second, it reconfigures memory encoding pathways by diminishing retrieval practice essential for long-term consolidation. Third, it alters deliberative reasoning by fostering automation bias, where users defer to AI outputs without engaging in independent critical evaluation. Understanding these mechanisms is essential for designing interventions that preserve cognitive autonomy while harnessing AI's benefits.

Theoretical foundations of cognitive offloading

Cognitive offloading is rooted in extended mind theory (Clark & Chalmers, 1998), which proposes that external tools, such as notepads, calculators, or smartphones, function as extensions of internal cognition. Historically, offloading reduced mental burden and freed cognitive resources for higher-order reasoning. For example, writing allows humans to externalise information, thus enhancing complex planning.

However, the emergence of AI represents a qualitative leap. Unlike static tools, AI dynamically processes, interprets, and generates content with semantic coherence and aesthetic polish, often surpassing the user's

unassisted outputs. This raises critical questions about how reliance on AI affects cognitive development, skill acquisition, and memory consolidation.

AI as a new cognitive partner

Recent studies suggest that AI transforms traditional cognitive offloading into delegated thinking, where users not only store information externally but also adopt AI-generated outputs with minimal critical scrutiny. For instance, students using AI to write essays may bypass the cognitive struggle involved in structuring arguments and synthesising sources, potentially impeding deep learning.

To illustrate this cognitive reconfiguration, consider a graduate student researching climate policy. Traditionally, the student would read primary sources, identify key arguments, mentally organize themes, and iteratively draft synthesis paragraphs—a process engaging sustained attention, working memory, and elaborative encoding. With AI assistance, the student instead inputs a prompt such as "summarise arguments for carbon taxation" and receives a coherent summary within seconds. The cognitive architecture shifts fundamentally: attention moves from deep engagement with source material to evaluating AI output quality; memory encoding bypasses the effortful retrieval practice that consolidates learning; and deliberative reasoning contracts to acceptance or rejection of pre-formed arguments rather than their independent construction.

Moreover, in professional contexts, AI-driven decision support systems influence managerial and financial choices. While such systems enhance efficiency, they risk diminishing analytical rigour if users uncritically adopt AI recommendations (Jarrahi, 2018).

Neural pathways and memory formation

The human brain relies on active retrieval and elaboration to consolidate memories into long-term storage (Karpicke & Roediger, 2008). Cognitive offloading short-circuits this process. Sparrow et al. (2011) demonstrated that knowing information is stored externally reduces the likelihood of encoding

it internally. The phenomenon is known as the Google Effect. AI amplifies this effect by not only storing but generating content, allowing users to skip the generation and integration stages crucial for memory and skill development.

Furthermore, neuroplasticity implies that repeated cognitive habits shape neural pathways (Draganski et al., 2004). Thus, chronic reliance on AI for cognitive tasks may rewire the brain away from active recall and critical evaluation toward superficial skimming and passive acceptance.

Implications for critical thinking and problem-solving

Critical thinking involves analysing, evaluating, and synthesising information to reach reasoned judgments. AI tools often provide well-structured, grammatically polished, and seemingly authoritative outputs. However, their internal logic may not always be sound, especially when based on probabilistic pattern recognition rather than grounded understanding (Bender et al., 2021).

If users adopt AI outputs without scrutinising underlying assumptions, argument structures, and evidence quality, the practice of critical thinking atrophies. This risk is particularly acute in education, where student reliance on AI undermines pedagogical objectives of cultivating independent analytical capacity.

Creativity and cognitive offloading

Creativity involves the recombination of disparate ideas into novel configurations. Offloading brainstorming to AI, such as generative text or image models, can stimulate new connections and accelerate ideation. However, if AI becomes the primary source of creative input, users may become curators rather than creators, selecting among AI outputs rather than generating ideas from personal insight and internalised knowledge (Carr, 2011).

This phenomenon resembles the difference between reading to expand imagination and relying exclusively on summaries, thereby limiting

depth. AI as a creativity tool is double-edged: it amplifies potential but risks fostering intellectual passivity if uncritically used.

Societal and educational implications

At a societal level, pervasive cognitive offloading to AI could reshape intellectual autonomy—the capacity to think independently and resist manipulation. If AI providers centralise cognitive scaffolding, the locus of epistemic authority shifts from individuals and distributed expert communities to proprietary algorithms (Zuboff, 2019).

In education, teachers face dilemmas on integrating AI constructively while ensuring students engage in authentic cognitive effort. Blanket bans may deprive students of essential digital literacy, while unregulated use risks hollowing out core competencies in writing, problem-solving, and critical appraisal (Selwyn et al., 2018).

Ethical and cultural dimensions

Ethically, the delegation of cognitive tasks to opaque AI systems raises concerns about epistemic opacity (Humphreys, 2009). Users cannot fully understand the basis of AI outputs due to their complexity and proprietary nature. This limits accountability and informed critique.

Culturally, cognitive offloading alters conceptions of expertise, creativity, and authorship. When AI generates art, literature, or analysis, questions arise: Who is the author? Who owns the ideas? What does it mean to think creatively when machines generate endless variations with minimal input? These questions challenge traditional intellectual property frameworks and cultural norms around human exceptionalism and creative authenticity (Coeckelbergh, 2022).

Strategies for sustainable cognitive offloading

To mitigate risks while harnessing AI's benefits, several strategies are proposed:

Metacognitive training: Educating users to reflect critically on when and how to offload tasks, maintaining a balance between external delegation and internal cognitive engagement (Risko & Gilbert, 2016).

Transparency in AI outputs: Developing AI systems that provide explanatory rationales rather than black-box answers, enhancing user understanding and trust calibration (Doshi-Velez & Kim, 2017).

Educational integration frameworks: Designing curricula that integrate AI as a tool for enhancing, rather than replacing, cognitive processes. For example, using AI to generate counterarguments that students must refute.

Policy and ethical oversight: Establishing regulatory and institutional frameworks to ensure AI systems are designed and deployed in ways that support rather than undermine human cognitive autonomy (Floridi & Cows, 2019).

Methodology and theoretical framework

This study adopts an interdisciplinary methodology combining cognitive science, educational psychology, and human-computer interaction to examine the effects of AI-driven cognitive offloading on human cognition. Grounded in Extended Mind Theory (Clark & Chalmers, 1998), which argues that external tools such as notebooks or calculators extend cognitive processes, the study investigates how AI tools further transform this by not only storing but dynamically generating content. This shift is conceptualised as delegated thinking, marking a qualitative departure from traditional cognitive offloading.

Key theoretical foundations include the Google Effect (Sparrow et al., 2011), where reliance on digital storage reduces internal memory encoding. The study extends this to AI, suggesting that generative AI tools may entirely bypass retrieval practice, undermining memory consolidation. Furthermore, automation bias (Logg et al., 2019) is examined, highlighting

the tendency to overtrust algorithmic outputs, a risk particularly salient in AI-integrated sectors such as education (essay generation) and healthcare (diagnostic support).

The methodological approach consists of literature synthesis, reviewing peer-reviewed studies on cognitive offloading, AI's cognitive impacts, and automation bias, along with case study analysis of Hatt University's AI integration, illustrating sector-specific risks such as distortions in relative grading due to widespread AI use, alongside mitigation strategies including metacognitive training to preserve cognitive skills.

Results

The provided tables synthesise theoretical and empirical insights on cognitive offloading and artificial intelligence (AI), highlighting how traditional external cognitive aids differ from AI-driven tools, as well as their implications across education, healthcare, and creative domains.

Table 1 contrasts traditional cognitive offloading tools, such as notebooks or calendars, with AI-based tools like ChatGPT and Midjourney. Traditional tools serve primarily as static storage that externalises information, thereby reducing immediate cognitive load while still requiring active retrieval (Sparrow et al., 2011). In contrast, AI tools are dynamic, generating, interpreting, and evaluating content. This shift transforms offloading from passive externalisation to delegated thinking, where AI not only stores information but produces and structures it for users (Sundar, 2020).

Table 2 outlines the risks and benefits of AI cognitive offloading in educational contexts. AI's capacity to free working memory for higher-order thinking is a significant benefit, allowing students to focus on synthesis and evaluation rather than rote memorisation. However, this can lead to superficial learning, as students bypass retrieval practice crucial for memory consolidation.

Table 3 explores sectors affected by AI cognitive offloading, focusing on automation bias. In education, AI essay writing tools promote efficiency but risk diminishing argument construction skills. In healthcare, AI diagnostic systems can improve accuracy, but Goddard, Roudsari, and Wyatt (2012) identify risks when clinicians uncritically accept AI recommendations.

Table 4 proposes strategies to mitigate negative effects while harnessing AI's benefits. Metacognitive interventions, such as training users to evaluate when and how to offload cognitive tasks, have been shown to improve strategic cognitive offloading. Explainable AI design, wherein algorithms provide transparent rationales for their outputs, can reduce automation bias by enhancing user understanding (Sundar, 2020).

Case study: Hatt University (Fictitious)

In 2024, Hatt University integrated AI writing assistants across undergraduate programs to enhance student productivity. Tools such as ChatGPT were embedded into the learning management system, enabling students to generate drafts, summarise readings, and refine grammar seamlessly. The policy aimed to improve digital literacy and reduce cognitive load in routine academic tasks.

Sara, a second-year psychology major, initially used AI to refine grammar and structure her essays. However, as semester demands increased, she began using AI to draft entire essays. Under Hatt University's relative (norm-referenced) grading system, student grades are distributed along a bell curve to maintain grade quotas. This meant that when more students began using AI tools, the baseline essay quality shifted upwards artificially.

The relative grading system exacerbated these effects. Instead of assessing absolute mastery of learning outcomes, grades reflected AI-assisted output quality. As more students used AI for drafting, the threshold for a passing or competitive grade rose. This created systemic pressure for even initially resistant students to adopt AI writing tools to avoid grade disadvantage, regardless of whether it hindered their learning.

Recognising these risks, Hatt University revised its AI integration policies in 2025, adopting metacognitive training workshops, explainable AI outputs, active learning integration, and piloting a criterion-referenced grading model, aligning grades to absolute learning outcomes rather than relative distribution.

Conclusion

The pervasive integration of artificial intelligence (AI) into cognitive tasks has transformed how humans think, learn, and create. This article has explored the phenomenon of cognitive offloading in the AI era, highlighting its dual role as both a facilitator of efficiency and a potential disruptor of deep cognitive engagement. While AI tools like ChatGPT and Midjourney offer unprecedented opportunities for productivity and creativity, their unchecked use risks undermining memory consolidation, critical thinking, and intellectual autonomy.

These findings carry significant normative implications across three domains. In education, institutions bear an ethical obligation to redesign curricula that require students to demonstrate independent cognitive engagement before accessing AI assistance. For reasoning practices, professionals in law, medicine, and finance should adopt structured protocols that mandate independent analysis prior to consulting AI recommendations. Regarding human–AI interaction design, developers should implement friction mechanisms—such as mandatory reflection prompts or delayed AI responses—that encourage users to engage cognitively before receiving AI outputs.

AI-driven cognitive offloading is not inherently detrimental; its impact depends on how society chooses to integrate and regulate these tools. By adopting proactive policies—grounded in metacognition, transparency, and ethical oversight—we can ensure AI serves as a cognitive enhancer rather than a crutch. The goal is not to resist technological progress but to shape it in ways that amplify human potential while safeguarding the intellectual rigor and creativity that define us.

References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Carr, N. (2011). *The shallows: What the Internet is doing to our brains*. W. W. Norton & Company.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>
- Coeckelbergh, M. (2022). *AI ethics*. The MIT Press. <https://doi.org/10.7551/mitpress/12549.001.0001>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. <https://doi.org/10.48550/arXiv.1702.08608>
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature*, 427(6972), 311–312. <https://doi.org/10.1038/427311a>
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>
- Humphreys, P. (2009). The philosophical novelty of computer simulation. *Synthese*, 169(3), 615–626. <https://doi.org/10.1007/s11229-008-9435-2>
- Jarrah, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. <https://doi.org/10.1016/j.bushor.2018.03.007>
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>

- Selwyn, N., Nemorin, S., Bulfin, S., & Johnson, N. F. (2018). *Everyday schooling in the digital age: High school, high tech?* Routledge.
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776–778. <https://doi.org/10.1126/science.1207745>
- Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74–88. <https://doi.org/10.1093/jcmc/zmz026>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Appendix: Tables

Table 1

Cognitive Offloading: Traditional Tools vs. AI Tools

Dimension	Traditional Cognitive Offloading Tools	AI-Based Cognitive Offloading Tools
Nature of Tool	Static storage (e.g. notebooks, calendars)	Dynamic generative systems (e.g. ChatGPT, Midjourney)
Function	Externalise and store information	Generate, interpret, and evaluate information
Impact on Memory	Reduces memory load; retrieval still required	May bypass memory encoding and retrieval entirely
Impact on Critical Thinking	Encourages planning and structured thought	Risks adoption of outputs without analysis (automation bias)
Creative Potential	Provides raw storage for later ideation	Generates novel combinations but may foster passive selection
Example References	Sparrow et al. (2011)	Sundar (2020); Logg et al. (2019)

Note. Own elaboration.

Table 2

Risks and Benefits of AI Cognitive Offloading in Education

Category	Potential Benefits	Potential Risks
Memory Formation	Frees capacity for higher-order synthesis	Reduces retrieval practice and deep encoding
Critical Thinking	Provides counterarguments or alternative explanations	Encourages surface learning and uncritical adoption
Creativity	Stimulates divergent thinking via novel AI outputs	May replace internal ideation processes
Skill Development	Enhances productivity and digital literacy	Undermines writing, analysis, and problem-solving skills
Example References	Sundar (2020)	Logg et al. (2019)

Note. Own elaboration.

Table 3

Cognitive Offloading and Automation Bias: Sectors Affected

Sector	AI Cognitive Offloading Example	Risk of Automation Bias
Education	AI essay writing assistance	High: Over-reliance on AI-generated arguments
Healthcare	AI diagnostic tools	High: Clinicians may accept AI errors uncritically
Finance	AI financial advising	Medium: Overtrust in algorithmic risk assessment
Creative Industries	AI art and music generation	Medium: Passive curation replaces active creation

Note. Own elaboration.

Table 4

Proposed Strategies for Mitigating Negative Effects of AI Cognitive Offloading

Strategy	Description	Expected Outcome
Metacognitive Interventions	Training users to reflect on when and how to offload tasks	Enhances strategic, balanced offloading
Explainable AI Design	AI systems provide transparent rationale for outputs	Reduces automation bias; fosters critique
Active Learning Integration	Embedding AI outputs into exercises requiring human modification or evaluation	Preserves cognitive engagement

Note. Own elaboration.

Inteligencia Artificial y erosión de los vínculos comunitarios: Una lectura filosófica desde el pensamiento de Byung-Chul Han

Artificial Intelligence and the erosion of community ties:
A philosophical reading based on the thought of Byung-Chul Han

Boris Osvaldo Saavedra Pérez

bsaavedrap@docente.uss.cl

Universidad San Sebastián

ORCID ID 0000-0003-3307-907X

Resumen: El presente artículo propone una lectura crítica, desde la perspectiva filosófica de Byung-Chul Han, acerca de las principales repercusiones que la IA y la digitalización han generado en la forma de vida del individuo de la modernidad tardía. A este respecto, Han advierte que la digitalización no solo optimiza procesos, sino que también contribuye a promover un régimen de la positividad, que excluye la negatividad constitutiva del ser humano.

Con este propósito, el artículo se estructura en torno a tres núcleos temáticos. En primer lugar, aborda el vaciamiento del sujeto, entendido como la pérdida de la interioridad reflexiva, la sensibilidad afectiva y la capacidad de construir sentido. En segundo lugar, examina la erosión de la comunidad, la cual es desplazada por formas de interconexión digital que, lejos de fomentar la relación simbólica y la pertenencia, generan una colectividad fragmentada y autorreferencial, sin densidad ni narrativa común. Finalmente, se plantea la posibilidad de una reapropiación ética de la *humanitas*, basada en prácticas que resisten el paradigma tecnocrático, como el silencio, la demora, la inactividad, la contemplación, la fiesta y la religión. Estas experiencias, marginadas por la lógica del rendimiento y la transparencia, permiten

vislumbrar otro horizonte para la existencia humana: uno que recupere la alteridad, el misterio y el espesor simbólico de la vida compartida.

Palabras clave: Inteligencia artificial; erosión; comunidad; filosofía; Byung-Chul Han.

Abstract: This article offers a critical reading, from the philosophical perspective of Byung-Chul Han, of the major repercussions that AI and digitalization have generated in the way of life of the individual in late modernity. In this regard, Han warns that digitalization not only optimizes processes but also contributes to promoting a regime of positivity that excludes the constitutive negativity of the human being.

With this aim, the article is organized around three thematic cores. First, it examines the hollowing out of the subject, understood as the loss of reflective interiority, affective sensitivity, and the capacity to generate meaning. Second, it analyzes the erosion of community, which is displaced by forms of digital interconnection that, far from fostering symbolic relations and belonging, produce a fragmented and self-referential collectivity lacking density and a shared narrative. Finally, the article considers the possibility of an ethical reappropriation of humanitas, grounded in practices that resist the technocratic paradigm, such as silence, delay, idleness, contemplation, festivity, and religion. These experiences, marginalized by the logic of performance and transparency, open the prospect of an alternative horizon for human existence, one that restores alterity, mystery, and the symbolic depth of shared life.

Keywords: Artificial intelligence; erosion; community; philosophy; Byung-Chul Han.

Introducción

La novedad suscitada por la inteligencia artificial (IA) no solo constituye un fenómeno tecnológico o científico, sino que también ha devenido en uno de los principales vectores de transformación de las estructuras simbólicas, sociales y antropológicas que han sustentado históricamente la experiencia humana. En efecto, la aceleración del desarrollo tecnológico, lejos de ser un simple avance instrumental, configura una nueva racionalidad que afecta de manera decisiva la forma en que el sujeto se comprende a sí mismo, se relaciona con los otros y se inserta en la comunidad. Considerando este escenario, el pensamiento del filósofo Byung-Chul Han ofrece una crítica incisiva y pertinente al impacto que la digitalización, la automatización y el dominio algorítmico han provocado, tanto en la configuración de la subjetividad contemporánea como en la posibilidad misma de una vida comunitaria significativa.

Desde la visión de Han, la IA se presenta como un dispositivo paradigmático de la *psicopolítica*, que contribuye mediante su utilización a suprimir la negatividad inherente a la experiencia humana, provocando que el dolor, la angustia, la alteridad y el misterio sean vistos como disfunciones a ser erradicadas. Siguiendo esta lógica, la IA no se limita solo a resolver problemas, más bien, redefine la acción humana por medio una óptica funcionalista y utilitaria; en la que lo ineficiente, lo improductivo o lo incalculable queda fuera del horizonte de sentido.

Así, uno de los aspectos más preocupantes que el filósofo surcoreano-alemán identifica es la disolución de la subjetividad como centro de experiencia y reflexión. En tanto que, en el régimen digital, el sujeto ya no es comprendido como un agente ético y narrativo, sino como un medio de rendimiento que se autoexplota y se adapta a un flujo constante de datos y estímulos. Sin embargo, la afectación no se limita al sujeto aislado. La IA también puede influir radicalmente en las formas de vinculación entre los seres humanos, erosionando la base afectiva, simbólica y ética sobre la que se construye la vida en comunidad.

De ahí que la comunidad, entendida como un espacio de donación, hospitalidad y narración compartida, sea sustituida por redes de conexión. Donde la interacción es mediada por plataformas que instrumentalizan el vínculo, a través de la lógica del algoritmo, lo que implica eliminar el riesgo y la apertura inherente a la relación con el otro. En este contexto, el otro ya no aparece como rostro que interpela, o como alteridad que desestabiliza; al contrario, aparece como un dato que puede ser procesado, clasificado y reformulado.

Considerando lo anterior, este artículo propone una lectura crítica, desde el pensamiento filosófico de Byung-Chul Han, acerca de las principales repercusiones que la IA y la digitalización han generado en la forma de vida actual. Con ello, no se pretende caer en una demonización simplista de la tecnología, al contrario, se intenta comprender las condiciones estructurales que hacen posible su dominio actual y preguntarse por sus efectos sobre la experiencia comunitaria.

Por lo tanto, el análisis se estructurará en torno a tres núcleos fundamentales: El vaciamiento del sujeto, la erosión de la comunidad, y la posibilidad de una resistencia ética basada en la reapropiación de la *humanitas*. Asimismo, se invita, a partir del pensamiento de Han, a repensar la relación entre técnica y espíritu. Para ello, se requiere un nuevo imaginario que supere la lógica de la eficiencia y el control, con la finalidad de restituir las dimensiones olvidadas de la existencia humana.

La IA y el vaciamiento del sujeto

Uno de los aspectos más provocadores del pensamiento de Byung-Chul Han es el examen que realiza sobre la noción de “inteligencia artificial”, pues, para el filósofo surcoreano-alemán, la inteligencia no puede reducirse solo a la capacidad de procesar datos o resolver problemas, lo que implica dimensiones existenciales irreductibles a la lógica algorítmica. De este modo, Han (2023a) sostiene:

La inteligencia artificial no puede pensar desde el momento en que no está capacitada para el *páthos*. El sufrimiento y el padecimiento son estados que no pueden ser alcanzados por máquina alguna. A las máquinas les es ajena, sobre todo, la inactividad contemplativa. Solo conocen dos estados: encendido y apagado. (p. 51)

En efecto, la inteligencia humana se distingue por su capacidad reflexiva, su apertura a la negatividad, la demora y la contemplación de las cosas. En cambio, la IA opera sobre la base de cálculos estadísticos, correlaciones y patrones, sin comprensión ni conciencia. En consideración a esto, Han (2022) plantea: “La inteligencia artificial no razona, sino que computa” (p. 59).

Así, aunque la IA se encuentre muy avanzada, aún no ha logrado desarrollar un razonamiento filosófico o dialéctico, por lo cual sigue operando mediante datos que le son proporcionados. En otras palabras, la IA no piensa, en realidad procesa. “La inteligencia artificial no piensa porque nunca está *fuera de sí misma*. El *espíritu* originariamente está *fuera de sí mismo* o *estremecido*. La inteligencia artificial puede *calcular* con rapidez, pero le falta el *espíritu*” (Han, 2021b, p. 54).

Esta distinción es crucial, porque, si bien la IA puede simular el pensamiento y la toma de decisiones, carece de la profundidad hermenéutica que define al espíritu humano. Computar no es narrar, calcular no es comprender. La narración implica una orientación hacia el sentido, una relación temporal, ética y afectiva con el mundo, algo que trasciende la lógica instrumental de la IA. De acuerdo con esto, Han (2023b) advierte: “La inteligencia artificial se maneja sin conceptos. *La inteligencia no es espíritu*. Solo el espíritu es capaz de establecer un nuevo orden de cosas, de desarrollar una nueva narración. La inteligencia calcula y computa. *Pero el espíritu narra*” (pp. 82-83).

Entonces, a diferencia de la IA que opera sobre correlaciones estadísticas y patrones repetitivos, el espíritu humano crea discontinuidad, introduce rupturas y posibilita el surgimiento de lo nuevo. Donde la IA optimiza lo existente, el espíritu humano interrumpe, interpreta y resignifica.

De ahí que, confiar ciegamente en la IA para orientar la existencia humana resulta ilusorio y conlleva la renuncia de aquello que constituye el núcleo ético y simbólico de la propia humanidad: la capacidad de dialogar, de reflexionar y de transformar el mundo a través de la palabra.

Desde otra perspectiva, resulta especialmente significativo cómo Han le otorga al dolor un rol fundamental en la diferenciación que realiza de la “inteligencia humana” y la “inteligencia artificial”. Precisamente, a través de sus ensayos, insiste en que el dolor constituye una experiencia irreductible a cualquier programación y, a su vez, lo señala como un elemento constitutivo del pensamiento auténticamente humano. Esto lo expresa enfáticamente al sostener:

El dolor es lo que distingue al pensar del calcular, de la inteligencia artificial. Inteligencia significa ‘escoger entre’ (inter-legere). Es una capacidad de discernimiento. Por eso no se sale del ámbito de lo ya existente. No es capaz de engendrar lo totalmente distinto. En eso se diferencia del espíritu. El dolor hace que el pensamiento sea más profundo. (Han, 2021a, pp. 62-63)

Con esta afirmación, el filósofo surcoreano-alemán subraya que la auténtica “inteligencia” es aquella que eleva el *espíritu*, lo cual requiere atravesar por la pasión y el dolor, algo que las máquinas; por definición, no pueden experimentar. “El *pathos* es el comienzo del pensamiento. La inteligencia artificial es *apática*, es decir, sin *pathos*, sin *pasión*. Solo *calcula*” (Han, 2021b, p. 56).

De esta manera, el pensamiento humano se distingue de forma radical, principalmente, porque embarga todo aquello que es afectivo. Tal como señala el filósofo surcoreano-alemán:

Lo *afectivo* es esencial para el pensamiento humano. *La primera afectación del pensamiento es la carne de gallina*. La inteligencia artificial no puede pensar porque no se le pone la carne de gallina. Le falta la dimensión afectivo-analógica, la emoción que los datos y la información no

pueden comportar. El pensamiento parte de una *totalidad* que precede a los conceptos, las ideas y la información. Se mueve ya en un «*campo de experiencia*» *antes* de dirigirse específicamente a los objetos y los hechos que encuentra en él. (Han, 2021b, p. 53)

Como resultado, la IA carece de esa dimensión afectiva que provoca la conmoción emocional, la cual no puede ser reducida a datos ni procesada mediante la información¹. Por eso, el pensamiento según Byung-Chul Han no parte de conceptos aislados ni de información cuantificable, sino de una totalidad previa que lo orienta, un *campo de experiencia* que antecede y fundamenta cualquier operación cognitiva. Esta dimensión pre-reflexiva, que involucra al cuerpo, al tiempo y al mundo compartido, es precisamente lo que queda excluido de la lógica funcional de la IA.

Sumado a lo anterior, Han (2021a) reafirma que la IA no es más que una racionalidad instrumental sin *espíritu*, una forma vacía de operar:

La inteligencia artificial no es más que un aparato de cálculo. Es capaz de aprender, incluso es capaz de *deep learning* o «aprendizaje profundo», pero es incapaz de tener experiencias. *El dolor es lo único que transforma la inteligencia en espíritu*. No hay *algoritmos del dolor*. (p. 63)

Ahora bien, esta crítica trasciende el plano meramente técnico y pone de manifiesto un trasfondo ético-antropológico más profundo, ya que la negación del dolor se ha convertido en uno de los ideales promulgados en la sociedad contemporánea. En concordancia con esta idea, Valle Jiménez (2019) señala que la cultura digital tiende a invisibilizar el sufrimiento, la infelicidad, las fragilidades y las disfunciones de orden psíquico, promoviendo en su lugar un ideal de optimización personal que responde a las exigencias de eficiencia del propio sistema. Sin embargo, esta búsqueda incesante de

¹ Lo propuesto por Han, trae a la memoria la obra literaria escrita por Philip K. Dick (2019) titulada ¿Sueñan los androides con ovejas eléctricas? Ya que, en ella, el protagonista llamado Rick tiende a reflexionar sobre la razón por la cual los androides fracasan ante una prueba de empatía, concluyendo que dicha capacidad, constituye un vínculo exclusivamente humano, algo que no puede ser simulado ni replicado artificialmente, pues se funda en una apertura ética y afectiva hacia el otro.

superación y rendimiento termina por volverse contra el sujeto, llevándolo al agotamiento, a hábitos autodestructivos e incluso al suicidio como expresión extrema de una subjetividad colapsada.

Este mandato de la autoexigencia permanente ha sido igualmente abordado en estudios previos, como Saavedra (2024b), quien señala:

Por ello, para el filósofo surcoreano, el sujeto se ha vuelto un empresario de sí mismo, puesto que está obsesionado con el rendimiento, lo que da lugar a numerosas enfermedades neuronales, como la depresión, el trastorno por déficit de atención con hiperactividad (TDAH), el trastorno límite de la personalidad (TLP) o el síndrome de desgaste ocupacional (SDO), entre otras. (p. 97)

Conforme a lo descrito, el sujeto neoliberal no se enfrenta a un enemigo externo, sino que él mismo se convierte en su propio enemigo. Siendo verdugo y víctima de una libertad malentendida, que se presenta como autoexplotación.

Así, surge uno de los ejes fundamentales de la teoría de Han, la *psicopolítica*. La cual sostiene que el poder contemporáneo ya no se ejerce mediante la represión o la prohibición, al contrario, ahora es la seducción, la optimización del rendimiento y la autoexplotación, que se convierten en los nuevos estándares que someten al individuo de la modernidad tardía.

Siguiendo esta lógica, la IA y la digitalización se integran dentro de los mecanismos de control por medio de los cuales opera la *psicopolítica*. Esto permite la manipulación del sujeto, debido a la acumulación y al procesamiento masivo de los datos o, lo que el filósofo denomina como *Big Data*:

El *Big Data* es un instrumento *psicopolítico* muy eficiente que permite adquirir un conocimiento integral de la dinámica inherente a la sociedad de la comunicación. Se trata de un *conocimiento de dominación* que permite intervenir en la psique y condicionarla a un nivel prerreflexivo. (Han, 2014b, p. 25)

Pero lo más preocupante, como advierten Niño, Cabeza y Flórez (2024), es que esta manipulación no se impone violentamente, sino que es más bien consentida y facilitada por el propio sujeto:

El primer tema abordado por Han para desarrollar su caracterización de la psicopolítica será la libertad, que deberá ser entendida de un manera distinta en el momento en que el individuo se constituye como un proyecto de sí mismo y que se acompaña con la sensación de libertad (Han, 2014, p. 7), puesto que el hombre considera que ya no está sujeto a elementos externos, sin embargo, ahora es él quien se somete a sí mismo en búsqueda de una mejora continua, en consecuencia, la existencia del sujeto se desarrolla en una sociedad de rendimiento. (p. 254)

En relación con esto, el sujeto al concebirse como un proyecto de sí mismo, termina interiorizando los mecanismos de control propios de la *psicopolítica*. La búsqueda de la autorrealización se convierte así en un mandato autoimpuesto que orienta su conducta y regula su vida cotidiana. Bajo esta lógica, el individuo es inducido a evitar la inactividad, el dolor o cualquier experiencia negativa, ya que estas prácticas son interpretadas como obstáculos para el rendimiento y la productividad constante.

Así pues, el sujeto es gradualmente reconfigurado como una entidad transparente, predecible y funcional, ajustándose a los parámetros del tecnocapitalismo. De este proceso emerge lo que Byung-Chul Han denomina como *Homo digitalis*:

El *homo digitalis* es cualquier cosa menos nadie. Él mantiene su identidad privada, aun cuando se presente como parte del enjambre. En efecto, se manifiesta de manera anónima, pero por lo regular tiene un perfil y trabaja incesantemente para optimizarlo. En lugar de ser *nadie*, es un *alguien penetrante*, que se expone y solicita la atención. (Han, 2014a, p. 28)

Este nuevo individuo, conceptualizado por Byung-Chul Han, se configura como un “alguien” cuya existencia depende de la constante

exposición de sí mismo. Lejos de ser una figura pasiva, se convierte en el ente ideal de la *psicopolítica*, pues no se le exige obedecer mandatos externos, sino autoproducirse y visibilizar cada aspecto de su vida, como parte de una “economía de la atención”².

En este contexto, la intimidad, el silencio y el misterio pierden legitimidad, desplazados por una autoescenificación permanente, donde el valor del sujeto se calcula según su capital simbólico, su eficacia afectiva y su capacidad para generar visibilidad en un entorno digital hipercompetitivo y saturado de estímulos.

La desaparición de la alteridad y la erosión de la comunidad

Según lo expuesto anteriormente, Byung-Chul Han advierte que el uso excesivo de la IA y la digitalización no solo vacían al sujeto, igualmente debilitan los vínculos comunitarios, al sustituir las relaciones interpersonales por meras interacciones instrumentales, caracterizadas por la aceleración y la desmaterialización.

De acuerdo con esto, Han (2020) sostiene: “La digitalización debilita el vínculo comunitario por cuanto tiene un efecto descorporizante” (p. 23). Por lo que la comunidad deja de ser un espacio de encuentro genuino y pertenencia simbólica para devenir en una red de individuos autorreferenciales, centrados en la gestión permanente de su visibilidad. En consecuencia, “*la digitalización es una anestesia*” (Han, 2021a, p. 52), en tanto neutraliza la intensidad afectiva y simbólica del encuentro humano, sustituyéndolo por una interrelación mediada por algoritmos.

² Sobre la “economía de la atención” es relevante mencionar que:

El concepto de economía de la atención fue forjado por Herbert Simon en 1969. Para Simon, la creciente abundancia de información en la esfera productiva genera una gradual escasez de atención humana. A mayor información disponible, mayor la escasez y mayor el valor de nuestra atención. Con la masificación de los medios de comunicación, en particular del internet, la noción de economía de la atención introducida en 1969 se ha vuelto una categoría central para el análisis económico contemporáneo (Celis, 2019, p. 42).

En esta misma línea, el filósofo surcoreano-alemán, profundiza su diagnóstico al afirmar que: “*la digitalización conduce forzosamente a la desaparición del otro*” (Han, 2024, p. 69). Esto, debido a que, en un entorno gobernado por la lógica de la transparencia, la inmediatez y la autoafirmación, no hay lugar para la alteridad. De este modo, el otro deja de ser un *fin en sí mismo* para convertirse en un medio, una especie de elemento funcional dentro del circuito de la autoexposición y del entramado digital.

Por otro lado, desde la perspectiva de Byung-Chul Han, la utilización de la tecnología digital le resta predominio a la manera en que el sujeto se relaciona con la realidad, el cuerpo y la memoria. Efectivamente, el filósofo surcoreano-alemán asevera: “la digitalización desmaterializa y descorporeiza el mundo. También suprime los recuerdos. En lugar de guardar recuerdos, almacenamos inmensas cantidades de datos” (Han, 2021b, p. 10). Con esto, se advierte que el mundo digital sustituye progresivamente la experiencia vivida por una acumulación masiva de información, sin espesor existencial ni carga afectiva. En efecto, “*La propia realidad asume ya la forma de información y de dato. Se informatiza y se transforma en datos. Percibimos la realidad sobre todo como si fuera un conjunto de informaciones, la percibimos mediante informaciones*” (Han, 2023b, p. 25).

En función de lo anterior, la realidad se desmaterializa, ocasionando su reemplazo por meras representaciones digitales, que terminan siendo moldeables y manipulables. Resultando evidente la semejanza con lo enunciado por el sociólogo Zygmunt Bauman (2003) quien describe este fenómeno como una forma de liquidez, donde las estructuras sólidas y duraderas de la vida social, se disuelven en una volatilidad constante, sin arraigo ni permanencia sustancial.

En relación con esto, Aguirre y Casas (2022), al comentar la postura crítica de Byung-Chul Han frente al proceso de digitalización, sostienen que:

Para él los medios digitales, el big data y la inteligencia artificial, además de suponer un potentísimo instrumento coercitivo para la libertad y actuación del ser humano (con el narcisismo como palanca fundamental en el marco de la psicopolítica) empobrecen el

pensamiento y el ejercicio filosófico, limitando las posibilidades del ser humano cara al futuro. (pp. 49-50)

Lo que facilita entender cómo, bajo la digitalización contemporánea, el sujeto es instrumentalizado y reducido funcionalmente, siendo parte de un sistema de vigilancia constante, en el que no solo es observado desde fuera, sino que, además, valida el hecho de vigilarse a sí mismo. Sobre esto, Byung-Chul Han subraya que el individuo digitalizado se convierte en una suerte de panóptico interiorizado:

En cuanto sujeto que se ilumina y vigila a sí mismo, está aislado en un panóptico en el que es simultáneamente recluso y guardián. El sujeto en red, digitalizado, es un *panóptico de sí mismo*. Así pues, se delega a cada uno la vigilancia. (Han, 2014b, p. 93)

Esta forma de autovigilancia permanente tiene efectos devastadores en la constitución de la subjetividad, debido a que, según el diagnóstico de Byung-Chul Han, el sujeto de la modernidad tardía ha ido perdiendo progresivamente su *humanitas*, al quedar expuesto en un entorno tecnificado que lo fragmenta, lo cuantifica y lo transforma en una entidad funcional.

Así, en lugar de irse formando a partir de experiencias significativas, conflictos o narrativas personales, el yo se diluye en la fluidez de datos, métricas y retroalimentaciones instantáneas. Como Han (2016) afirma: “Hoy, el yo es muy pobre en cuanto a formas de expresión estables con las que pudiera identificarse y que le otorgaran una identidad firme. Hoy nada tiene consistencia” (p. 26). Esta fragilidad del yo no puede entenderse únicamente como un déficit interno, sino que también es el resultado de la desaparición de una alteridad auténtica, que confronta y delimita al sujeto. Asimismo, Peña Zerpa (2023) advierte:

Si existe algo determinante en la construcción del yo es la existencia de una alteridad auténtica, no consumible, no pulida. Es decir, el yo se compone por una relación dialéctica entre el sí mismo y la

existencia del otro que no necesariamente es igual y parecido a mí mismo. (p. 61)

En consideración a lo descrito, ambas perspectivas convergen en una crítica fundamental; sin la fricción del otro, sin un límite simbólico que ponga en cuestión la propia imagen, el yo se vacía de consistencia y se convierte en una función adaptativa dentro del ecosistema digital.

Precisamente, debido a la pérdida de consistencia y cohesión social ocurre lo que Han denomina *enjambre*, una forma de colectividad propia de la era digital que se encarga de sustituir a la comunidad tradicional. Al respecto, Almeyda (2022) expresa:

Sumado a lo dicho, la sociedad del enjambre digital elimina la ley del nómós, la cual posibilita, debido a su negatividad, un entendimiento del mundo en que se habita. En este marco de ideas, el pensamiento es dejado de lado por el cálculo haciendo que la negatividad propia de la búsqueda por la verdad, que viene de la mano del pensar, sea reemplazada por la positividad del procedimiento aditivo de lo medible y lo comparable. Todo esto se vuelve determinante para pensar la existencia que se vive dentro de la sociedad del rendimiento. (pp. 300-301)

Conforme a esto, el *enjambre* carece de unidad simbólica, de narración histórica y de voluntad colectiva, simplemente porque no se configura como una vinculación política ni como una comunidad orientada con fines compartidos. Al contrario, se trata de un conjunto de individualidades, conectadas pero desvinculadas entre sí. Por ello, Han (2014a) examina estas cualidades y señala: “Los individuos que se unen en un enjambre digital no desarrollan ningún *nosotros*. Este no se distingue por ninguna concordancia que consolide la multitud en una masa que sea sujeto de acción” (p. 27).

Con ello, el filósofo surcoreano-alemán evidencia cómo la comunidad tradicional tiende a disolverse en el entorno digital, dando paso a formas de interacción cada vez más efímeras y despersonalizadas. De manera que, en lugar de un diálogo auténtico, predominan las reacciones instantáneas

mediadas por símbolos gráficos que sustituyen la comunicación interpersonal.

Complementando lo anterior, Almeyda (2022) argumenta: “La medialidad digital, dice Han, destruye la distancia entre la esfera de los asuntos públicos y privados en un solo lugar sobrepuesto que no permite una distinción adecuada entre las personas” (pp. 298-299). Por lo tanto, la interacción se limita a una circulación de imágenes y mensajes que privilegian la inmediatez por sobre la profundidad, dando lugar a una visibilidad vacía, donde todo se expone por medio de la pantalla. Sin embargo, como se ha hecho hincapié, se carece absolutamente de un verdadero encuentro real. Esto se enmarca en la lógica neoliberal, que da origen a las comunidades digitales, o *community*, una forma de asociación digital que, como parte del *enjambre*, se rige por intereses funcionales y no por vínculos reales. En este contexto, Byung-Chul Han manifiesta que:

La comercialización y mercantilización total de la cultura ha tenido por efecto la destrucción de la comunidad. La *community* que tan a menudo invocan las plataformas digitales es una forma de comunidad mercantil. La comunidad como mercancía es el fin de la comunidad. (Han, 2012b, p. 31)

Este argumento se intensifica cuando el filósofo surcoreano-alemán aborda el discurso dominante en torno a la *community*, debido a que ésta se encuentra regida por los parámetros económicos del mercado. Asimismo, lejos de propiciar un espacio de comunión o pertenencia real, este modelo de comunidad digital se revela como una construcción ilusoria que encubre el aislamiento creciente de los individuos. Sobre esto, el filósofo señala: “Hoy se evoca por todas partes la *community*, pero esta es una forma mercantil de comunidad. No permite que surja ningún *nosotros*. El consumo desatado aísla y aleja a las personas. Los consumidores están solos” (Han, 2023a, p. 14).

Entonces, la promesa de comunidad que ofrecen las plataformas digitales encubre, en realidad, una profunda fragmentación del tejido social, donde el sujeto es reducido a consumidor y los vínculos comunitarios a una

interfaz funcional. Siguiendo esta misma idea, Romero Orduña (2024) sostiene:

Según Han, en medio de las cotidianas prácticas digitales hipercomunicativas e hiperexpresivas subsumidas en los controles psicopolíticos contemporáneos, se termina descorporalizando a los sujetos, despersonalizando y expandiendo la comunicación y entablando interconexiones sin mirada ni voz pero dominadas por un silencio ensordecedor. (p. 168)

Así, lo que se configura en la era digital no es una auténtica comunidad, sino una forma de aislamiento sistemático y crónico, donde la presencia del otro ya no se sustenta en el reconocimiento recíproco. Lo que provoca que sea filtrada y mediada por algoritmos que determinan la visibilidad, el alcance y la naturaleza misma de la interacción.

En definitiva, Byung-Chul Han expone que la digitalización no solo compromete la comunicación, sino que también participa de manera irreversible en la erosión de los vínculos comunitarios. Por lo tanto, la comunidad, entendida como un espacio de colaboración, escucha activa y construcción simbólica compartida; es sustituida por burbujas digitales que reducen al otro a un mero reflejo del propio yo, neutralizando la alteridad y eliminando la posibilidad de un encuentro genuino.

Hacia una reapropiación ética de la humanitas

En los apartados anteriores se ha evidenciado el diagnóstico radicalmente crítico de Byung-Chul Han respecto al impacto de la IA y la digitalización. Sin embargo, es fundamental advertir que su postura no debe ser reducida a una simple tecnofobia. Por el contrario, en el trasfondo de su pensamiento se vislumbra la necesidad urgente de una reapropiación ética de la *humanitas* frente al avasallador progreso de la tecnología digital. Esta reapropiación no implica rechazar el progreso tecnológico, sino más bien reinscribirlo dentro de un horizonte existencial que recupere la centralidad de la alteridad, la experiencia del encuentro y la densidad simbólica de un mundo verdaderamente compartido.

Así, la primera condición para llevar a cabo esta reapropiación ética de la *humanitas* que se infiere de la obra de Han, consiste en restituir aquellas dimensiones existenciales que el orden digital ha relegado al margen, tales como: el silencio, la demora, la inactividad y la contemplación. En vista de que, lejos de constituir meras formas de pasividad improductiva, estos modos de *estar-en-el-mundo* revelan una profundidad olvidada por causa de la hiperactividad y el rendimiento permanente.

En este contexto, Han (2020) destaca que: “El silencio hace estar a la escucha. Va acompañado de una peculiar receptividad, de una profunda atención contemplativa. La actual presión para comunicar conduce a que no podamos cerrar los ojos ni la boca. Esa presión profana la vida” (p. 54). Por tal razón, el silencio se concibe como la condición de posibilidad tanto para el encuentro consigo mismo, como para el encuentro con el otro y el mundo. En consonancia con esto, Byung-Chul Han recupera la tradición clásica de la *theoría* como una forma de contemplación que otorga sentido duradero a la existencia. Según afirma:

La mayor felicidad brota del demorarse contemplativo en la belleza, antiguamente llamada *theoria*. Su sentido temporal es la duración. Se ocupa de las cosas eternas e inmutables, que descansan en sí mismas. Ni la virtud ni la sabiduría, solo la entrega contemplativa a la verdad acerca al hombre a los dioses. (Han, 2015, p. 125)

Esta entrega a lo contemplativo se contrapone rotundamente al paradigma acelerado y fragmentario de la digitalización, donde el tiempo ya no se experimenta como duración significativa, sino como una sucesión de instantes cuantificables y consumibles. Por lo que recuperar la *theoría* implica, entonces, restaurar una relación no instrumental con la realidad, en la que el valor de las cosas no se encuentre mediado exclusivamente por su utilidad o su rendimiento. Asimismo, Han (2023a) sintetiza con claridad lo propuesto cuando señala que: “La inactividad es una *forma de esplendor* de la existencia humana” (p. 11).

Con ello, sugiere que la inactividad no debe confundirse con la pereza ni con la apatía, al contrario, pues se trata de una forma de presencia plena, que resiste la lógica de la productividad y devuelve al tiempo su espesor vital. Es por esto que el mismo Han (2023a) enfatiza:

El callar le da profundidad al habla. Sin silencio no hay música, sino nada más que ruido y alboroto. El juego es la esencia de la belleza. Allí donde solo reina el esquema de estímulo y reacción, necesidad y satisfacción, problema y solución, propósito y acción, la vida degenera en supervivencia, en desnuda vida animal. La vida solo recibe su resplandor de la inactividad. (pp. 12-13)

Ahora bien, la propuesta de la inactividad da paso a la contemplación, la que se alza como un contrapeso necesario frente al imperativo del hacer constante. Porque el rendimiento sin interrupción reduce la vida a mera subsistencia funcional, desprovista de verdadero significado. Así, “La contemplación se contrapone a la producción” (Han, 2023a, p. 52). De ahí que el filósofo surcoreano-alemán reivindique una revolución silenciosa basada en el no-hacer, el demorarse y la amabilidad³, como modos de resistencia frente al rendimiento y la hiperconectividad.

De igual manera, este giro hacia la inactividad como forma de revitalización del tiempo ha sido explorado en una investigación anterior, destacando su potencial existencial:

En consecuencia, la inactividad constituye el ideal de la vida humana, pues promueve un tiempo verdaderamente libre, donde se persigue la auténtica realización vital, lo que implica volver a tener acceso a la

³ Es interesante lo señalado por Almeyda respecto a la amabilidad, en tanto que recoge y amplía el planteamiento de Han, quien la concibe como una forma de negatividad en abierta oposición al régimen neoliberal. Según dice: “Han desarrolla la amabilidad como un modo negativo de demora que genera interrupción temporal al concebir su objeto no como consumo, sino como experiencia de la nueva praxis temporal no distorsionada ni atomizada” (Almeyda, 2022, p. 311).

esencia de la realidad humana y circundante, que se encuentra fuera del orden del trabajo y la producción. (Saavedra, 2024a, p. 245)

Por otro lado, las condiciones precisadas para hacer frente al paradigma tecnocapitalista se integran perfectamente en “la metáfora del jardín”⁴, que viene a representar una imagen ética de la comunidad no violenta, hospitalaria y atenta:

La ética del jardín busca reestructurar los vínculos humanos que se han desgastado por la aceleración actual que el sistema promueve para mantener un control sobre el enjambre del rendimiento. La filosofía de Han, desde esta interpretación, no busca iniciar una revolución de lo público en el sentido moderno, sino que se ubica en el plano de una revolución privada del sujeto, esto último implica ubicarse en la dimensión de las prácticas íntimo-cotidianas, para retejer esos lazos sociales desde una ética oriental que haga posible hablar entre personas más allá de la psicopolítica del sistema. (Almeyda, 2023, pp. 117-118)

Así es como “la metáfora del jardín” remite a un espacio de cuidado, de diversidad y de armonía relacional, donde la tecnología no desaparece, pero sí se subordina a los ritmos orgánicos de la existencia. Con esto, entonces, lo que se pretende es someter la tecnología a principios éticos orientativos que trasciendan su finalidad técnica.

De igual manera, esta perspectiva encuentra eco en las reflexiones hechas por Llano Alonso (2024), quien subraya la necesidad de una base ética robusta para encaminar el desarrollo tecnológico:

⁴ Cabe señalar que “la metáfora del jardín” remite a una interpretación vivencial de la filosofía de Byung-Chul Han. En este contexto, su pensamiento adquiere una dimensión concreta y espiritual, pues el filósofo surcoreano-alemán encontró en el cultivo de su propio jardín una forma de revolución silenciosa en contra de los males que aquejan a la sociedad de la modernidad tardía. Esta experiencia fue plasmada por Han (2019) en su obra *Loa a la Tierra: Un viaje al jardín*, donde reivindica el contacto con la tierra, el ritmo natural y la atención cuidadosa como formas de resistencia frente a la digitalización y la desconexión del mundo.

En realidad, la ética es un ingrediente básico para que la investigación tecnológica pueda avanzar tranquilamente, explorar sus propias posibilidades y ser beneficiosa para la humanidad, en la medida que propicia el mantenimiento de un alto nivel de confianza entre los actores implicados en el proceso de la IA. (p. 188)

La confianza, en este marco, no es un mero efecto secundario del progreso tecnológico, antes bien, es un requisito para su legitimidad social. Porque solo una tecnología amparada en principios éticos sólidos puede sostener la promesa de servir al bien común, evitando su instrumentalización en beneficio de objetivos puramente mercantiles. En otras palabras, la tecnología debería estar al servicio de la comunidad y no promover su erosión, ya que solo así puede concebirse una comunidad sustentada en la escucha, la memoria compartida y la construcción simbólica del mundo.

Ahora bien, otro aspecto relevante en el pensamiento de Byung-Chul Han para una posible reapropiación ética de la *humanitas* frente a la tecnología, se encuentra en su reflexión sobre la fiesta y la religión⁵. Debido a que ambas prácticas, en su sentido más originario, constituyen formas de resistencia simbólica frente a la fragmentación, el aislamiento y el narcisismo extremo que caracterizan al individuo de la modernidad tardía.

Lejos de ser simples momentos de evasión, la fiesta y la religión aparecen como espacios de comunión, donde se suspenden las actividades intrascendentes y se recupera una experiencia compartida del tiempo y del sentido. Como Han (2023a) señala: “La fiesta, por el contrario, crea comunidad. Reúne y une a las personas. El sentimiento de festividad es siempre un *sentimiento de comunidad*, un *sentimiento-de-nosotros*” (p. 72).

⁵ Es importante advertir que Han aborda la religión de manera general, aunque mayormente se centra en el cristianismo; aun así, le interesa dejar en claro que la religión es una forma de experiencia simbólica y comunitaria que permite resistir la lógica de la positividad, reconectando al sujeto con la contemplación, la alteridad y el misterio que excede a la racionalidad instrumental.

De ahí que la fiesta no implique solamente una instancia de celebración, igualmente cumple una función mucho más profunda, la de restituir los vínculos comunitarios y reactivar un sentimiento de pertenencia compartido. Así, en medio de un mundo predominantemente tecnificado y regido por lógicas de cálculo, eficiencia y consumo, la fiesta se presenta, para Han, como una forma originaria de comunidad. Se trata de un acontecimiento simbólico que suspende la fragmentación y genera aquel *sentimiento-de-nosotros* capaz de resistir el aislamiento estructural de la modernidad tardía.

Asimismo, esta potencia comunitaria de la fiesta se vincula estrechamente con su enraizamiento en lo narrativo y lo religioso. En este sentido, Byung-Chul Han sostiene que la religión, y particularmente el cristianismo, ofrece una metanarración que confiere unidad y sentido a la totalidad de la existencia:

La religión es una narración característica, con una verdad intrínseca. Con su manera de *narrar*, nos *salva* de la contingencia. La religión cristiana es una metanarración, que no omite ningún rincón de la vida y la fundamenta en el ser. El tiempo mismo cobra un sentido narrativo. El calendario cristiano hace que cada día tenga su sentido. En la era posnarrativa, el calendario pierde su carácter narrativo y se convierte en una agenda vaciada de sentido. Las festividades religiosas son los clímax y los apogeos de una narración. Sin narración no hay fiesta ni tiempo festivo, no hay sentimiento de festividad, vivida como una intensa sensación de ser; no hay más que trabajo y tiempo libre, producción y consumo. (Han, 2023b, p. 12)

De lo dicho, se desprende que es a través de su estructura ritual y simbólica, expresada en el calendario litúrgico, las festividades y los ciclos sagrados del tiempo; que la religión logra salvar al sujeto de la mera contingencia y la fragmentación temporal propia del mundo digital.

Por último, la fiesta y la religión, en tanto formas de suspensión del rendimiento, del cálculo y de la productividad, ofrecen encuentros significativos, temporalidades compartidas y vínculos no mediados por la

lógica tecnocapitalista. Estas experiencias, que resisten la compulsión de la hiperactividad y del rendimiento, revelan la posibilidad de otro horizonte en el camino a la reapropiación ética de la *humanitas*.

Conclusiones

En función de lo señalado, la crítica hacia la IA y la digitalización del filósofo surcoreano-alemán Byung-Chul Han, evidencia una preocupación profunda por las condiciones que hacen posible la experiencia humana significativa, como también por los vínculos simbólicos que configuran la vida en común. Puesto que, la IA irrumpe como un artefacto cultural que puede alterar de manera decisiva el pensamiento racional y auténticamente humano, sustituyéndolo por patrones de datos, y debilitando la experiencia afectiva mediante respuestas automatizadas.

Por lo tanto, el uso excesivo de la IA y de la tecnología digital termina generando un vaciamiento del sujeto; lo cual no constituye un efecto absoluto de los avances tecnológicos, sino más bien es el resultado estructural de una racionalidad algorítmica que se articula con la lógica neoliberal de la autoexplotación, la transparencia total y la vigilancia consentida. Ya que, tanto la IA como la digitalización forman parte de la *psicopolítica*, una nueva forma de poder que se encarga de someter al sujeto por medio de la estimulación productiva y el rendimiento sin límites. En este escenario, el individuo se define por su valor funcional dentro del sistema tecnocapitalista, lo que lo convierte en una entidad calculable, predecible y optimizable.

Asimismo, esta alteración subjetiva no puede entenderse de manera aislada, pues afecta directamente a los modos de vinculación y de configuración comunitaria. Por lo que el ideal de una comunidad basada en el reconocimiento recíproco, la comunicación significativa y la construcción simbólica compartida, acaba siendo reemplazada por el ascenso de nuevas formas de colectividad sin densidad, ni sustancialidad. Tal es el caso del *enjambre digital*, o la *community*, que no representan un “nosotros”, más bien, responden a una agregación dispersa de individualidades autorreferenciales, cuya interacción está mediada por algoritmos que privilegian la visibilidad sobre el vínculo, y la conexión sobre la relación.

En consecuencia, la comunidad tradicional se erosiona por un exceso de conectividad que neutraliza la presencia real del otro. Como lo ha advertido Han, en la era digital la alteridad desaparece, el otro se transforma en un espejo funcional del yo, y la distancia simbólica que permite la construcción de una subjetividad sólida queda abolida. Entonces, no es solo una crisis relacional, también es una crisis de sentido, pues sin un “otro real” que confronte, interpele y limite, el yo queda atrapado en la circularidad del reconocimiento inmediato, perdiendo profundidad, identidad y orientación.

No obstante, aunque el diagnóstico realizado por Byung-Chul Han destaque por mostrar un carácter sombrío, igualmente es interesante que exista la posibilidad de considerar sus advertencias para plantear una reapropiación ética de la *humanitas*, lo que exige modificar la lógica de la optimización para restituir los fundamentos simbólicos y existenciales de la vida compartida. Esta reapropiación de la *humanitas* no implica un rechazo absoluto de la tecnología, sino más bien, su reinscripción en una temporalidad distinta; la del silencio, la demora, la inactividad y la contemplación. Dichas categorías, que el filósofo recupera de la tradición clásica y religiosa, aparecen como respuestas frente a los ideales neoliberales de la modernidad tardía.

Así, en primer lugar, el silencio y la demora se erigen como actos de resistencia frente a la compulsión digital de la estimulación constante. En una época marcada por la saturación de estímulos y la hipercomunicación, el silencio, el demorarse y el detenerse permiten reconstruir la relación con el mundo, devolviendo a la experiencia su espesor simbólico. De igual manera, la inactividad y la contemplación provocan una ruptura con la lógica de la producción y del rendimiento, abriendo un espacio de resistencia frente a la colonización del tiempo por parte del sistema tecnocapitalista.

En segundo lugar, la fiesta y la religión, en tanto prácticas simbólicas originarias, representan formas de recomposición de los vínculos comunitarios. Frente a la fragmentación, el aislamiento y la eficiencia tecnificada, la festividad reconstituye el sentimiento de comunidad, suspendiendo temporalmente las jerarquías del rendimiento. Por lo tanto, la fiesta genera una reapertura al sentido, al otro y al tiempo compartido. A su

vez, la religión, provee un marco narrativo que otorga unidad a la existencia, estableciendo una metanarración que permite interpretar el devenir en clave de significado, más allá del cálculo y de la contingencia. Ambas prácticas devuelven al sujeto una orientación simbólica que lo reconecta con la comunidad, con la historia y con lo sagrado.

En definitiva, el desafío ético no reside en eliminar la IA ni en añorar una sociedad anti-tecnológica, sino en subordinar el desarrollo tecnológico a fines humanos y comunitarios, donde la IA lejos de constituirse como una amenaza ineludible, pueda ser repensada como una herramienta al servicio del cuidado, la reciprocidad y la hospitalidad. Para ello, se requiere un marco ético robusto que oriente su implementación, promoviendo una inteligencia sensible a la fragilidad de lo humano. En otras palabras, se trata de desplazar el centro de gravedad del progreso técnico desde la optimización hacia el sentido, desde el control hacia el reconocimiento, desde el cálculo hacia la narración.

Finalmente, resulta urgente pensar en una reapropiación ética de la *humanitas*, que oriente el desarrollo tecnológico en torno al florecimiento humano, y, a su vez, favorezca la formación de sujetos más receptivos, contemplativos y capaces de habitar un mundo en común.

Referencias

- Aguirre, A., & Casas, R. (2022). La complejidad antropológica de la comunicación digital y la inteligencia artificial y cómo afrontarla: Un diálogo entre Yuval Noah Harari, Byung-Chul Han y Paolo Benanti. *Revista Ciencias de la Complejidad*, 3(2), 46–64. <https://doi.org/10.48168/cc022022-004>
- Almeyda, J. (2022). Prácticas de la amabilidad: una interpretación del pensamiento de Byung-Chul Han. *Areté*, 34(2), 291–318. <https://doi.org/10.18800/arete.202202.001>
- Almeyda, J. (2023). *Hacia una ética del jardín: Estudios filosóficos sobre el pensamiento de Byung-Chul Han* (1.ª ed.). Ediciones UIS.
- Bauman, Z. (2003). *Modernidad líquida* (1.ª ed.). Fondo de Cultura Económica.
- Celis, C. (2019). Economía de la atención y visión maquínica: hacia una semiótica asignificante de la imagen. *Hipertextos*, 5(7), 41–55. <https://revistas.unlp.edu.ar/hipertextos/article/view/7759>

- Dick, P. K. (2019). *¿Sueñan los androides con ovejas eléctricas?* (7.ª ed.). Cátedra.
- Han, B.-C. (2014a). *En el enjambre* (1.ª ed.). Herder.
- Han, B.-C. (2014b). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder* (1.ª ed.). Herder.
- Han, B.-C. (2015). *El aroma del tiempo: Un ensayo filosófico sobre el arte de demorarse* (1.ª ed.). Herder.
- Han, B.-C. (2016). *La salvación de lo bello* (1.ª ed.). Herder.
- Han, B.-C. (2019). *Loa a la tierra: Un viaje al jardín* (1.ª ed.). Herder.
- Han, B.-C. (2020). *La desaparición de los rituales: Una topología del presente* (1.ª ed.). Herder.
- Han, B.-C. (2021a). *La sociedad paliativa: El dolor hoy* (1.ª ed.). Herder.
- Han, B.-C. (2021b). *No-cosas: Quiebras del mundo de hoy* (1.ª ed.). Herder.
- Han, B.-C. (2022). *Infocracia: La digitalización y la crisis de la democracia* (1.ª ed.). Herder.
- Han, B.-C. (2023a). *Vida contemplativa: Elogio de la inactividad* (1.ª ed.). Herder.
- Han, B.-C. (2023b). *La crisis de la narración* (1.ª ed.). Herder.
- Han, B.-C. (2024). *La tonalidad del pensamiento. Trilogía de las conferencias* (Vol. 1). Paidós.
- Llano Alonso, F. H. (2024). Ética(s) de la inteligencia artificial y Derecho: Consideraciones a propósito de los límites y la contención del desarrollo tecnológico. *Derechos y Libertades: Revista de Filosofía del Derecho y Derechos Humanos*, (51), 177–199. <https://doi.org/10.20318/dyl.2024.8587>
- Niño, E. E., Cabeza, O. J., & Flórez, 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>
- Peña Zerpa, C. A. (2023). Niveles de formación en Inteligencia Artificial (IA): entre la ola y su distanciamiento. *Temas De Comunicación*, (47), 50–63. <https://doi.org/10.62876/tc.vi47.6324>
- Romero Orduña, L. M. (2024). Byung-Chul Han: subjetividades contemporáneas y nuevas formas de dominación desde los espacios digitales. *BAJO EL VOLCÁN. REVISTA DEL POSGRADO DE SOCIOLOGÍA. BUAP*, 161–190. <https://doi.org/10.32399/ICSYH.bvbuap.2954-4300.2024.5.10.748>
- Saavedra, B. (2024a). Filosofía de la inactividad: sobre la noción de vida contemplativa en Tomás de Aquino y Byung-Chul Han. *Revista De Filosofía UCSC*, 23(1), 235–248. <https://doi.org/10.21703/2735-6353.2024.23.1.2327>
- Saavedra, B. (2024b). El cuerpo humano como mercancía pornográfica en *La sociedad de la transparencia*. *Metanoia: Revista académica de la Escuela Profesional de Filosofía de la Universidad Antonio Ruiz de Montoya*, 9(1), 92–111. <https://doi.org/10.53870/metanoia20241346>
- Valle Jiménez, D. (2019). Dolor y autoexplotación en la era digital. *Contrastes. Revista Internacional De Filosofía*, 23(3), 163–180. <https://doi.org/10.24310/Contrastescontrastes.v23i3.6599>

Modelo de alfabetización en inteligencia artificial como complemento a la alfabetización informativa

Artificial intelligence literacy model as a complement to
information literacy

Alejandro Villegas-Muro

avillegas@uach.mx

Universidad Autónoma de Chihuahua

ORCID ID 0000-0001-6362-5137

Resumen: La rápida adopción de sistemas de IA en educación superior está reconfigurando las prácticas de lectura, escritura y evaluación, y exige literacidades que trascienden la alfabetización informacional (ALFIN). Este artículo propone un modelo de Alfabetización en Inteligencia Artificial (AIL) orientado a universidades latinoamericanas y alineado con pedagogías de la lengua. Metodología: se realizó una meta-síntesis cualitativa de estudios recientes, con cribado temático, codificación apoyada en software y evaluación de calidad metodológica (CASP) para derivar categorías y relaciones. Resultados: emergen cuatro dimensiones integradas de AIL—técnico-conceptual (nociones, funcionamiento y límites de la IA), crítico-interpretativa (sesgos, trazabilidad y análisis del discurso mediado por algoritmos), aplicada-práctica (usos académicos responsables en lectura y escritura) y socio-ético-cívica (normativa, derechos y participación)—articuladas como una “doble hélice” con puntos de control de equidad. Se identifican el rol catalizador de las bibliotecas universitarias, la conveniencia de integración temprana en el currículo de Lengua y la necesidad de criterios de evaluación con transparencia algorítmica. Conclusiones: la AIL complementa y expande ALFIN, ofreciendo pautas para rediseñar planes de estudio, servicios bibliotecarios y políticas institucionales; aporta instrumentos para fortalecer la lectura crítica, la escritura académica con trazabilidad y la formación ética frente a riesgos de la IA, contribuyendo a reducir brechas digitales y de poder informacional.

Palabras clave: Alfabetización en inteligencia artificial; alfabetización informacional; pedagogías de la lengua; educación superior; bibliotecas universitarias.

Abstract: The rapid adoption of AI systems in higher education is reshaping reading, writing, and assessment practices, and demands literacies that transcend information literacy (IL). This article proposes an Artificial Intelligence Literacy (AIL) model tailored to Latin American universities and aligned with language pedagogies. Methodology: a qualitative meta-synthesis of recent studies was conducted, including thematic screening, software-assisted coding, and methodological quality appraisal (CASP) to derive categories and relationships. Results: four integrated AIL dimensions emerge—technical-conceptual (AI notions, functioning, and limits), critical-interpretive (biases, traceability, and analysis of algorithm-mediated discourse), applied-practical (responsible academic uses in reading and writing), and socio-ethical-civic (regulation, rights, and participation)—articulated as a “double helix” with equity checkpoints. The study identifies the catalytic role of university libraries, the benefits of early integration into the Language curriculum, and the need for assessment criteria with algorithmic transparency. Conclusions: AIL complements and expands IL, offering guidelines to redesign curricula, library services, and institutional policies; it provides tools to strengthen critical reading, academic writing with traceability, and ethical formation in the face of AI risks, helping to reduce digital and informational power gaps.

Keywords: Artificial intelligence literacy; information literacy; language pedagogies; higher education; university libraries.

Introducción

La inteligencia artificial (IA) ha dejado de ser un dominio reservado a la investigación especializada para convertirse en la “capa invisible” que modula transacciones financieras, diagnósticos médicos, filtros de información y decisiones administrativas en instituciones públicas y privadas por igual. En

el mundo existe la creciente brecha entre el entusiasmo normativo y la preparación efectiva de la ciudadanía sigue siendo notable. La evidencia internacional demuestra que, cuando los individuos carecen de competencias para interrogar los procesos algorítmicos, las desigualdades socioeconómicas preexistentes se ensanchan (Wang et al., 2024). Este fenómeno, conocido como “AI divide”, no solo reproduce la brecha digital clásica (relativa al acceso físico a la tecnología), sino que introduce una división cognitiva y ética: ¿quién comprende los sesgos de entrenamiento, las limitaciones de los modelos, las implicaciones legales de la automatización y las responsabilidades derivadas de su uso?

En el ámbito académico, las universidades mexicanas enfrentan un reto doble. Por un lado, deben capitalizar las ventajas pedagógicas de herramientas generativas que prometen personalizar retroalimentación, automatizar rúbricas o potenciar la producción científica; por otro lado, han de salvaguardar la autenticidad, la creatividad y la integridad académica (Coman & Cardon, 2024; Van Niekerk et al., 2024). El problema se agrava cuando dichas herramientas son empleadas sin un acompañamiento crítico que permita a docentes y estudiantes reconocer sus limitaciones epistémicas: la opacidad de la “caja negra” algorítmica, los sesgos de entrenamiento y la tendencia a naturalizar su autoridad (Domínguez y Stoyanovich, 2023). En este contexto, la alfabetización informativa resulta necesaria, pero ya no suficiente. Se impone la construcción de un modelo de Alfabetización en Inteligencia Artificial (AIL) que complemente, sin sustituir, los programas de Alfabetización Informativa (ALFIN) vigentes.

A nivel global, organismos como la IFLA abogan por que las bibliotecas universitarias se erijan en nodos formativos para la AIL, subrayando la necesidad de ofrecer servicios de referencia, talleres y objetos de aprendizaje que incorporen perspectivas técnicas y socio-éticas (IFLA, 2024). Sin embargo, estudios diagnósticos señalan que los propios bibliotecarios aún requieren una capacitación especializada para mediar entre los sistemas algorítmicos y las comunidades académicas (Watkins & Johnson, 2024). Esta tensión refleja una paradoja: los espacios institucionales mejor posicionados para liderar la alfabetización en IA son, al mismo tiempo, los

que exhiben lagunas formativas relevantes. A ello se suma la constatación de que la alfabetización en datos (entendida como la habilidad para interpretar y gestionar grandes volúmenes de datos) constituye la piedra angular de la alfabetización en IA (Ghodoosi et al., 2024; Schüller, 2022). Ignorar la dimensión datificada conduciría a un enfoque meramente instrumental que reproduce la lógica de “usuarios pasivos” sin capacidad de agencia crítica.

Asimismo, los marcos regulatorios emergentes, como la propuesta de AI Act de la Unión Europea, muestran un giro hacia la gestión de riesgos y la rendición de cuentas algorítmica (Sachoulidou, 2024). Aunque en algunos países de latinoamérica no han adoptado un instrumento legal equivalente, la convergencia normativa internacional configura un horizonte en el que las universidades deberán acreditar el dominio de conceptos como trazabilidad, explicabilidad y auditoría de algoritmos. Desde la didáctica, revisiones sistemáticas confirman que la integración temprana y progresiva de la IA en los planes de estudio favorecen tanto el aprendizaje conceptual como el desarrollo de competencias éticas (O’Dea et al., 2024). Sin embargo, la literatura advierte que las iniciativas aisladas resultan insuficientes frente a la magnitud del desafío; se requiere un enfoque transversal e interdisciplinario que abarque todas las áreas del conocimiento y fomente la colaboración entre docentes, bibliotecarios, tecnólogos y responsables de políticas institucionales (Lo, 2023).

En este escenario, el presente trabajo se articula como un metaanálisis cualitativo que sintetiza 31 estudios empíricos y teóricos para responder a la pregunta orientadora: ¿Cómo puede articularse un modelo de Alfabetización en Inteligencia Artificial, en el contexto latino, que complemente y potencie la Alfabetización Informativa universitaria? La elección de la técnica metaanalítica obedece a dos razones. Primero, la producción científica sobre IA y educación crece a un ritmo exponencial, lo que genera hallazgos disgregados y, a veces, contradictorios. El metaanálisis permite identificar patrones convergentes, lagunas de conocimiento y factores contextuales que moderan los resultados. Segundo, al limitar la base documental a fuentes ya incluidas en la investigación original, garantizamos la coherencia bibliográfica

y evitamos la incorporación de referencias externas que pudiesen desalinearse de los objetivos iniciales.

La relevancia práctica de este ejercicio para las universidades es múltiple. En términos curriculares, los hallazgos ofrecerán directrices para diseñar asignaturas troncales, diplomados o certificaciones que integren competencias técnicas con competencias críticas. En el ámbito de los servicios de información, la propuesta delinearé estrategias para que las redes bibliotecarias adopten sistemas de recomendación éticas, desarrollen capacitaciones sobre búsquedas aumentadas por IA y establezcan protocolos de transparencia. Finalmente, en el plano de la gestión institucional, el modelo servirá como marco de referencia para políticas de innovación que eviten la adopción acrítica de plataformas algorítmicas, fomenten la ética de datos y garanticen la inclusión de poblaciones tradicionalmente subrepresentadas (Karan & Angadi, 2023).

Conviene subrayar que el propósito no es suplantar la alfabetización informativa, sino ampliar su perímetro epistemológico. Tal como advierten Flierl y Maybee (2020), la evolución de ALFIN hacia versiones críticas ya prefiguraba la necesidad de interrogar las relaciones de poder que moldean la producción y circulación de conocimiento. La irrupción de la IA sólo intensifica esta exigencia y obliga a desplazar el foco desde la fiabilidad de las fuentes hacia la opacidad del algoritmo. Así, la alfabetización en IA no se concibe como un apéndice temático, sino como un eje transversal que redefine el contrato pedagógico entre universidades y sociedad.

Entonces, la urgencia de construir capacidades ciudadanas para una convivencia responsable con la IA interpela de manera directa a las universidades, depositarias históricas de la formación de profesionales y agentes de cambio. Este metaanálisis pretende ofrecer un mapa conceptual y estratégico que, fundado en la evidencia existente, oriente la acción académica hacia una alfabetización integral que no solamente habilite el uso competente de algoritmos, sino que también cultive la agencia crítica necesaria para cuestionarlos y transformarlos cuando vulneren principios de equidad y justicia social.

Marco teórico

De la alfabetización informativa a la alfabetización en inteligencia artificial

La alfabetización informativa clásica (ALFIN) emergió a finales del siglo XX para dotar al estudiantado universitario de capacidades orientadas a encontrar, evaluar y utilizar información de manera ética. Su objetivo primordial era generar personas críticas en un ecosistema dominado por la abundancia de fuentes escritas y digitales. Con el tiempo, las corrientes críticas de ALFIN incorporaron una lectura sociopolítica que obliga a cuestionar los sistemas de poder que intervienen en la producción del conocimiento (Flierl & Maybee, 2020). Bajo esta lógica, ya no basta con identificar si un texto es fiable o si un autor tiene prestigio; es imperativo desentrañar quién decide qué información se legitima y qué voces son silenciadas.

La irrupción de la inteligencia artificial profundiza estas inquietudes. Modelos de aprendizaje automático clasifican, filtran y priorizan contenidos mediante procesos opacos y se alimentan de bases de datos que reproducen sesgos estructurales ligados a género, etnia o clase social (Domínguez & Stoyanovich, 2023). La ALFIN, esta centrada en la fiabilidad de la fuente, se confronta ahora con la necesidad de interpretar probabilidades, métricas de desempeño y matrices de confusión. De ahí que la comunidad académica proponga ampliar el constructo hacia una Alfabetización en Inteligencia Artificial (AIL) que contemple tres ejes inseparables: comprensión técnica básica de los sistemas, análisis socio-ético de sus implicaciones y empoderamiento cívico para participar en la toma de decisiones sobre su adopción (Schüller, 2022).

En el entorno universitario mexicano, esta evolución implica repensar planes de estudio y servicios bibliotecarios. No basta con añadir un módulo optativo sobre “nuevas tecnologías”; es preciso transversalizar la discusión sobre sesgos algorítmicos en cursos de métodos de investigación, derecho digital, comunicación científica e, incluso, en asignaturas de ética profesional. Tal integración responde a una lógica de alfabetización crítica que busca formar profesionales capaces de interpelar la lógica comercial y política de la

IA, en lugar de convertirse en consumidores acríticos de soluciones empaquetadas.

La alfabetización en datos como piedra angular

Las tecnologías de IA contemporáneas aprenden de conjuntos masivos de datos; por ello, cualquier déficit en alfabetización de datos se traduce directamente en una carencia de alfabetización en IA (Ghodoosi et al., 2024). Las competencias para recopilar, limpiar, visualizar e interpretar datos constituyen el sustrato sobre el cual se erige la comprensión de modelos predictivos y generativos. No sorprende, entonces, que los estudios bibliométricos registren un incremento exponencial en la investigación sobre “data literacy” desde 2015, reflejando un reconocimiento global de su importancia estratégica (Yan et al., 2024).

En el caso de México, donde persisten brechas significativas en habilidades estadísticas y de programación, la urgencia de fortalecer la alfabetización en datos es doble. Primero, porque habilita a estudiantes y docentes para evaluar la calidad de los conjuntos de entrenamiento que subyacen a sistemas de IA; segundo, porque fomenta la cultura de la evidencia, es decir, la capacidad de sustentar argumentos académicos y decisiones profesionales en análisis riguroso y reproducible.

En términos curriculares, esto se traduce en incorporar prácticas de minería de datos, ética de datos y visualización crítica dentro de los programas de licenciatura y posgrado, independientemente del área disciplinar. Desde las bibliotecas universitarias, la alfabetización en datos se refuerza mediante talleres sobre gestión de conjuntos abiertos, principios FAIR (Por sus siglas en inglés: Encontrables, Accesibles, Interoperables y Reusables) y uso ético de repositorios. Así, la alfabetización en IA deja de ser una isla temática y se construye sobre un ecosistema de datos gobernado por principios de transparencia y reproducibilidad científica.

Inteligencia híbrida humano-IA

Jarrahi et al. (2022) proponen un marco tripartito para conceptualizar la relación entre la inteligencia humana y la artificial: a) la inteligencia “más allá de lo humano”, donde el sistema opera de forma autónoma y supera la capacidad humana en tareas específicas; b) la inteligencia “dependiente del humano”, donde la máquina requiere supervisión continua; y c) la inteligencia “híbrida”, que combina fortalezas humanas (creatividad contextual, juicio ético con las ventajas algorítmicas, velocidad de procesamiento, análisis de grandes volúmenes de datos). Este enfoque subraya que el diseño sinérgico, y no la sustitución, debe guiar los objetivos educativos.

En el ámbito de la comunicación profesional, los arreglos híbridos permiten mantener estándares de claridad y coherencia, sin sacrificar la voz auténtica de la persona que escribe, siempre que exista transparencia sobre el rol de la IA en el proceso (Coman & Cardon, 2024). Para la universidad, el paradigma híbrido adquiere particular relevancia porque contempla la coautoría entre estudiantes y agentes algorítmicos: chatbots que facilitan borradores, sistemas de retroalimentación automática que sugieren mejoras estilísticas y plataformas de análisis que identifican patrones en bases de datos científicas. La clave radica en evitar la tentación de delegar el juicio crítico a la máquina y, en cambio, posicionar al estudiantado como “curador” y “verificador” del output algorítmico.

Adoptar este modelo conlleva desafíos éticos y pedagógicos. En primer lugar, se requiere desarrollar criterios de atribución autoral que distingan entre asistencia mecanizada y creación intelectual, cuestión ya discutida en políticas editoriales y normativas de integridad académica. En segundo término, los docentes necesitan formación especializada para diseñar actividades que exploten la complementariedad sin fomentar la pereza cognitiva. Finalmente, las instituciones deben establecer infraestructuras seguras y transparentes que permitan auditar los algoritmos empleados, en consonancia con la tendencia regulatoria internacional hacia la explicabilidad y la rendición de cuentas (Sachoulidou, 2024).

Por lo tanto, para que la ALFIN sea un complemento a la AIL supone reconocer que la información ya no es un objeto estático, sino un flujo mediado por sistemas capaces de aprender, clasificar y tomar decisiones. La alfabetización en datos opera como cimiento epistémico que habilita la lectura crítica de esos sistemas, mientras que el paradigma de la inteligencia híbrida ofrece un horizonte pedagógico donde la IA amplifica, pero no cancela, la agencia humana. Comprender esta tríada (alfabetización informativa crítica, alfabetización de datos y cooperación híbrida) resulta indispensable para las universidades latinas que aspiran a formar profesionales capaces de navegar, cuestionar y co-crear en un mundo crecientemente algoritmizado.

Metodología

Diseño del metaanálisis cualitativo

El presente estudio adopta un metaanálisis cualitativo cuyo propósito es articular, a partir de la evidencia ya existente, un marco de referencia para la alfabetización en inteligencia Artificial (AIL) en el ámbito universitario latinoamericano. En contraste con los metaanálisis estadísticos, la meta-síntesis cualitativa busca re-interpretar hallazgos heterogéneos para generar modelos conceptuales y líneas de acción. El proceso se desarrolló en cuatro fases recursivas: (1) delimitación del corpus, (2) extracción y sistematización de datos, (3) evaluación de calidad y credibilidad, y (4) síntesis inter-casos y generación de teoría de nivel medio.

Estrategia de búsqueda y selección de estudios

La base documental utilizada en este metaanálisis cualitativo procede de una búsqueda estructurada realizada en el marco de la investigación original sobre alfabetización en inteligencia artificial y alfabetización informativa. Dicha búsqueda tuvo como objetivo identificar estudios empíricos y teóricos recientes que abordaran, de manera explícita, competencias, prácticas educativas o marcos conceptuales relacionados con la inteligencia artificial, la alfabetización informativa, la alfabetización de datos y la ética algorítmica en contextos formativos.

Bases de datos y fuentes consultadas. La búsqueda de manuscritos se llevó las bases de datos Science Direct y SAGE. Esta combinación permitió captar tanto la discusión internacional sobre AI literacy como aportes procedentes de América Latina.

Palabras clave y ecuaciones de búsqueda. Se emplearon combinaciones booleanas de términos en inglés. A modo de ejemplo, una de las ecuaciones centrales fue:

(“artificial intelligence literacy” OR “AI literacy” OR “algorithmic literacy” OR “data literacy”) AND (“information literacy” OR “digital literacy”) AND (“higher education” OR university OR “academic libraries”).

Rango temporal. Se acotaron los resultados al periodo 2018–2024 para capturar la etapa de consolidación de la alfabetización en IA y de los debates en torno a la integración de sistemas generativos en educación superior. No obstante, se permitió la inclusión de trabajos previos cuando su contribución resultaba fundacional para la alfabetización informativa crítica o la alfabetización de datos y seguían siendo ampliamente citados en los estudios recientes.

Criterios de inclusión y exclusión. Se incluyeron artículos que cumplieran simultáneamente los siguientes criterios:

- a) Estar publicados en revistas académicas revisadas por pares o en informes profesionales de alta relevancia disciplinar;
- b) Abordar explícitamente al menos uno de los siguientes constructos: alfabetización en IA, alfabetización informativa, alfabetización de datos, literacidad crítica o ética algorítmica;
- c) Presentar un vínculo claro con contextos formativos (educación básica o superior, bibliotecas, formación profesional continua) y ofrecer hallazgos o propuestas con potencial de transferencia al ámbito universitario;

Se excluyeron documentos de divulgación sin aparato metodológico, literatura puramente técnica sobre desarrollo de algoritmos sin componente formativo, notas de opinión no analíticas y literatura gris no indexada (informes internos, blogs, presentaciones sin revisión externa).

A partir de esta estrategia se obtuvo un conjunto inicial de trabajos que, tras la lectura de títulos y resúmenes y la eliminación de duplicados, se redujo a 31 estudios considerados nucleares para la investigación original. Sobre ese corpus inicial se aplicaron posteriormente los criterios de pertinencia conceptual y transferibilidad educativa descritos en el subapartado “Delimitación del corpus”, lo que condujo al conjunto final de 29 estudios que nutren la meta-síntesis cualitativa.

Delimitación del corpus

Se partió de las 31 fuentes empíricas y teóricas ya citadas en la versión preliminar del manuscrito. Aunque el cuerpo documental originalmente abarcaba múltiples dominios (educación básica, educación superior, comunicación profesional, bibliotecología, salud y políticas públicas), se re-filtró con base en dos criterios de inclusión adicionales:

1. Pertinencia conceptual: el estudio debía abordar explícitamente competencias, prácticas o marcos teóricos relacionados con IA, alfabetización informativa, alfabetización de datos o ética algorítmica;
2. Transferibilidad educativa: los hallazgos debían ser susceptibles de adaptación a contextos formativos universitarios, aun si el estudio original se desarrolló en otro sector (por ejemplo, bibliotecas públicas o periodismo).

Veintinueve (29) artículos cumplieron plenamente ambos criterios; los dos restantes mantuvieron un estatus periférico y se utilizaron sólo para matizar discusiones específicas.

Extracción y sistematización de datos

Cada artículo se codificó manualmente mediante un protocolo de siete categorías, diseñadas ad hoc para la presente investigación:

- a) Dominio de aplicación (escuela básica, universidad, sector profesional, política pública, bibliotecas);
- b) Constructo de alfabetización principal (ALFIN, AA, alfabetización de datos, literacidad crítica, ética algorítmica);
- c) Población/Contexto (estudiantes, docentes, bibliotecarios, ciudadanos, decisores);
- d) Metodología (cuantitativa, cualitativa, mixta, revisión sistemática, conceptual);
- e) Hallazgos clave (variables significativas, barreras, facilitadores, recomendaciones);
- f) Nivel de granularidad (macro-político, meso-institucional, micro-aula);
- g) Potencial de transferencia (alto, medio, bajo).

Evaluación de calidad y credibilidad

Con el fin de ponderar la robustez de la evidencia, se aplicó una versión adaptada del Critical Appraisal Skills Programme (CASP) para estudios educativos y de políticas. Los criterios evaluaron: (1) solidez metodológica, (2) claridad en la descripción del contexto, (3) validez interna, y (4) relevancia para la alfabetización algorítmica. Ningún estudio fue excluido por calidad insuficiente; sin embargo, los puntajes CASP se emplearon como coeficiente de peso durante la síntesis, otorgando mayor influencia a los artículos con diseño riguroso (por ejemplo, las revisiones sistemáticas de Weng et al., 2024, y Yan et al., 2024).

Ética y reflexividad

Dado que el metaanálisis se basa en estudios previamente publicados, no involucró recolección de datos sensibles. No obstante, se mantuvo una postura reflexiva frente a los sesgos de interpretación propios de los investigadores, especialmente en lo relativo al contexto mexicano. Se reconocen dos limitaciones: (1) la exclusión de literatura posterior a 2024 podría dejar sin cubrir innovaciones recientes, y (2) la decisión de no incorporar fuentes externas limita la diversidad de perspectivas, pero refuerza la coherencia solicitada por el proyecto original.

En conjunto, esta metodología de meta-síntesis cualitativa no sólo reorganiza la evidencia existente, sino que la re-significa para diseñar un marco integral de alfabetización algorítmica aplicable a universidades mexicanas. Al priorizar la convergencia temática, la transferibilidad educativa y la credibilidad metodológica, el proceso garantiza que las recomendaciones emergentes estén sólidamente ancladas en datos y sean pertinentes para los actores institucionales responsables de formular políticas y prácticas de enseñanza-aprendizaje en torno a la inteligencia artificial.

Verificación bibliográfica preliminar

Antes de iniciar la síntesis se cotejó la lista de 31 referencias con todas las citas en el texto. El cotejo reveló que al menos dos referencias no poseían los criterios para la sección analítica; además, su foco (adopción de “chatbots” académicos e interpretación simultánea) se alejaba de los objetivos curriculares del presente estudio. Siguiendo la pauta de “coherencia corpus-citas”, ambos trabajos se excluyeron. El corpus final quedó en 29 estudios (ver Tabla 1).

Procedimiento de meta-síntesis cualitativa

1. Delimitación y cribado

Criterios de inclusión.

- Relación explícita con alfabetización informativa, de datos o en IA.
- Transferibilidad al entorno universitario (aula, biblioteca, política interna).

2. Codificación sistemática

Cada artículo se codificó en una base de datos manual con cinco ejes revisados y afinados tras el pre-test:

- a) Dominio de aplicación; b) Constructo principal de alfabetización;
- c) Población / grupo de interés; d) Diseño metodológico; e) Hallazgos nucleares.

3. Síntesis inter-casos

Mediante comparación constante se agruparon 119 códigos libres en 14 categorías temáticas; estas se refinaron hasta las cuatro dimensiones del modelo AIL (“doble hélice”). Se registró el proceso en un *audit-trail* accesible para réplica.

4. Aseguramiento ético y reflexivo

- Sin datos sensibles primarios → dispensa de comité ético.
- Reflexividad: los autores revisaron supuestos culturales (p. ej., latino-centrismo) en talleres de contraste con colegas europeos y asiáticos.

Tabla 1.

Tabla de metanálisis

Tabla del metanálisis					
No.	Referencia abreviada	Dominio / Contexto	Metodología	Hallazgo nuclear para la AIL	Potencial de transferencia *
1	Coman y Cardon 2024	Comunicación empresarial (HE)	Cuant. (n = 281)	Transparencia sobre co-autoría IA mantiene percepción de autenticidad	Alto
2	Cox y Mazumdar 2022	Bibliotecología	Conceptual	Define IA para bibliotecas; advierte ambigüedad terminológica	Alto
3	DeVasto y Palmer 2024	Formación empresarial (HE)	Estudio de caso	“Línea de preguntas críticas” para desmontar neutralidad algorítmica	Medio
4	Domínguez y Stoyanovich 2023	Política de datos	Marcos normativos	Propone modelo “stakeholder- first” para alfabetizar sobre sesgos	Alto
5	Flierl y Maybee 2020	Teoría ALFIN	Revisión teórica	Vincula alfabetización crítica con análisis de poder	Alto
6	Ghodoosi et al 2024	HE multipaís	Mixta	Carencia de “data literacy” limita adopción de IA	Alto
7	Hermann 2021	Medios personalizados	Teórico- crítico	Personalización extrema genera burbujas ideológicas	Medio
8	IFLA Journal 2024	Bibliotecas globales	Editorial	Bibliotecas como “catalizadores”	Alto

Tabla del metanálisis					
				de resiliencia tecnológica	
9	Jarvis et al 2024	Inglés profesional	Exp. quasi	Chatbots + feedback explicativo ↑ autoconfianza	Medio
10	Jarrahi et al 2022	Ciencia de la información	Ensayo	Tipología “más- allá / dependiente / híbrida” de IA	Alto
11	Karan y Angadi 2023	Educación escolar (India)	Revisión	Aboga por integración ética desde primaria	Medio
12	Lo 2023	Políticas de IA	Análisis comparado	Mapa de vacíos regulatorios en bibliotecas	Medio
13	O’Dea et al 2024	HE (UK y HK)	Cuant. (n = 447)	Coordinación institucional mejora literacia generativa	Alto
14	Pan et al 2024	Nuevos medios	Exp. laboratorio	Agencialidad percibida del bot afecta confianza	Medio
15	Piller 2024	Comunicación de crisis	Ensayo	Riesgos de desinformación por generación automática	Medio
16	Ponce 2024	RR.HH. / Empleabilidad	Exp. (n = 120)	Autoeficacia en <i>prompt</i> mejor a calidad de CV	Medio
17	Porlezza y Schapals 2024	Periodismo	Revisión	Falta de guías éticas en redacciones algorítmicas	Alto
18	Sachoulidu u 2024	Derecho tech (UE)	Jurídico	AI Act: marco de riesgos estratificados	Alto
19	Schüller 2022	Estadística pública	Ensayo	Propone “AI y Data literacy as civil right”	Alto
20	Su y Yang 2023a	Primaria	Diseño instruccional	“Cinco grandes ideas” para IA infantil + robótica	Medio

Tabla del metanálisis					
21	Su y Yang 2023b	HE global	Marco conceptual	Guía para integrar ChatGPT en docencia con ética	Alto
22	Van Niekerk et al 2024	Escritura académica	Revisión	Necesidad de directrices institucionales sobre IA	Medio
23	Vartiainen et al 2024	Primaria	Estudio cualit.	Sesgos en <i>text-to-image</i> como actividad reflexiva	Medio
24	Wang et al 2024	Opinión pública (UE)	Survey (n = 6.005)	Segmentos vulnerables = baja habilidad + escepticismo	Alto
25	Watkins y Johnson 2024	Bibliotecarios	Informe profesional	Falta de confianza para explicar IA a usuarios	Medio
26	Weng et al 2024	HE global	Revisión sist. (n = 120)	Éxito ↔ inmersión temprana + proyectos disciplinares	Alto
27	Wood et al 2021	Medicina	Survey	Alumnado exige IA clínica, docentes dudan preparación	Medio
28	Yan et al 2024	Bibliometría	Mapeo visual	Explosión de estudios de “data literacy” (2015-22)	Alto
29	Zou et al 2023	HE (China)	Mixta	Menor autoeficacia femenina en uso de generadores	Medio

Nota. Elaboración propia.

Resultados y análisis

Trayectorias pedagógicas

El metaanálisis revela un patrón que se repite desde la educación básica hasta los posgrados profesionales: la exposición temprana a conceptos de IA genera una trayectoria de aprendizaje sostenido que, con adecuadas reiteraciones, incrementa la complejidad cognitiva sin fracturar la motivación del estudiantado. En la etapa K-12, propuestas como las de Su y Yang (2023a) adaptan la robótica y el machine learning a marcos conceptuales basados en los “cinco grandes principios” del pensamiento computacional, lo que posibilita un puente temprano entre la manipulación tangible de sensores y la abstracción algorítmica. De modo complementario, Vartiainen et al. (2024) demuestran que introducir generadores de imágenes text-to-image en primaria favorece la comprensión de los sesgos de entrenamiento cuando los docentes guían discusiones metacognitivas sobre la procedencia de los datos.

En educación superior, la evidencia converge en la necesidad de integración transversal: O’Dea et al. (2024) documentan que la alfabetización generativa de IA en Reino Unido y Hong Kong mejora cuando las facultades coordinan políticas unificadas en lugar de relegar la cuestión a asignaturas optativas. Weng et al. (2024), mediante una revisión sistemática de 120 estudios, confirman que los programas universitarios más exitosos combinan inmersión temprana, refuerzo disciplinar y evaluación formativa basada en proyectos.

Los programas profesionales refuerzan la tendencia espiral al mostrar que la IA no solo optimiza prácticas clínicas o comunicativas, sino que consolida la identidad profesional cuando se integra críticamente. Wood et al. (2021) hallan, en medicina, que la aceptación de módulos sobre sistemas de apoyo al diagnóstico crece si el currículo primero aborda conceptos estadísticos básicos y, en fases posteriores, despliega simuladores de pacientes potenciados por IA. Jarvis et al. (2024) extienden la lógica al inglés profesional: los simuladores de entrevistas laborales incrementaron la autoconfianza de estudiantes cuando incluyeron sesiones de

retroalimentación explicativa, transparentando cómo el algoritmo calificaba la pronunciación, el contenido semántico y la expresión no verbal.

Un hallazgo transversal es la utilidad de metodologías activas: aprendizaje basado en problemas con conjuntos de datos reales, dramatizaciones con chatbots que ejemplifican sesgos y talleres de prompt engineering enfocados en la autoexplicación algorítmica. Estas estrategias no solo aumentan el compromiso, sino que fomentan la conciencia metacognitiva considerada predictor clave de transferibilidad profesional (Jarvis et al., 2024).

Imperativos socio-éticos

En todos los dominios analizados emergen con fuerza los dilemas de ética, equidad y sesgos (Domínguez & Stoyanovich, 2023; Hermann, 2021). El consenso radica en que la alfabetización algorítmica debe trascender la mera competencia técnica para desarrollar una sensibilidad crítica hacia la estructura de poder incrustada en los datos. Hermann (2021) advierte que la “personalización masiva” promete experiencias a la medida, pero, sin supervisión ética, deriva en burbujas informativas que perpetúan estereotipos. Piller (2024) traslada la discusión a la comunicación de crisis: los modelos generativos pueden amplificar o mitigar la desinformación, dependiendo de cómo se configure su entrenamiento y la transparencia de su uso.

Desde la perspectiva regulatoria, el AI Act europeo (Sachoulidou, 2024) introduce un enfoque de gestión de riesgos estratificado que obliga a las universidades a mapear en qué categoría encajan sus herramientas académicas y administrativas. Aunque algunos países de latinoamérica carecen de legislación equivalente, la asimetría regulatoria global crea el riesgo de “zonas francas” algorítmicas en las que las prácticas importadas ignoren salvaguardas esenciales (Lo, 2023). Las bibliotecas, como veremos en la siguiente subsección, se sitúan en el centro de esta tensión, pues manejan datos sensibles de comunidad y actúan como intermediarias entre proveedores de IA y usuarios finales.

En los ámbitos del periodismo y la gobernanza de la información, los estudios de Porlezza y Schapals (2024) evidencian que la automatización descontrolada erosiona la legitimidad democrática y la confianza pública. Consecuentemente, la alfabetización en IA debe incorporar nociones de responsabilidad compartida: no vale solo con formar usuarios críticos, sino profesionales capaces de diseñar y gestionar sistemas con principios de transparencia, explicabilidad y auditoría social.

La brecha de la IA

Un aporte distintivo de este metaanálisis es la conceptualización de la “brecha de la IA” como fenómeno multidimensional que combina alfabetización técnica, capital cultural y confianza en las instituciones. Wang et al. (2024), mediante modelado de clases latentes, identifican segmentos de la población que conjugan escepticismo y baja habilidad, lo que los hace particularmente vulnerables a decisiones automatizadas en salud, finanzas o justicia. En China, Zou et al. (2023) muestran que la adopción de herramientas generativas está mediada por el género, con mujeres que reportan menor autoeficacia en el diseño de prompts complejos, reflejando patrones de exclusión tecnológica históricamente documentados.

Para Latinoamérica, los países que poseen desigualdades digitales y educativas marcadas, este hallazgo es crucial: cualquier modelo de alfabetización que no integre puntos de control de equidad corre el riesgo de profundizar la brecha existente. Ello implica diseñar materiales plurilingües, flexibles y con ejemplos contextualizados —por ejemplo, datos abiertos sobre presupuesto municipal, que resuenen con la realidad local— e incluir mecanismos de acompañamiento para personas con discapacidades o rezagos de conectividad.

Bibliotecas y profesionales de la información como catalizadores

La Federación Internacional de Asociaciones de Bibliotecarios e Instituciones (IFLA, 2024) postula a las bibliotecas universitarias como hubs de alfabetización algorítmica, capaces de democratizar el acceso a herramientas de IA y de promover su uso responsable. Sin embargo, el metaanálisis revela

que tales aspiraciones chocan con déficits formativos de los propios bibliotecarios. Watkins y Johnson (2024) documentan que la mayoría se siente insegura al explicar conceptos como *overfitting* o interpretabilidad de modelos, mientras que Cox y Mazumdar (2022) subrayan la ambigüedad terminológica que aún rodea a “IA” en el gremio.

Para superar estas carencias, se recomienda un plan escalonado de *upskilling* que combine microcredenciales técnicas con espacios de reflexión ética y apropiación pedagógica. De igual forma, la biblioteca puede fungir como laboratorio vivo donde se prototipen buscadores semánticos, sistemas de recomendación y chatbots de referencia, siempre bajo protocolos de auditoría y con participación estudiantil. Así, el aula y la biblioteca se nutren mutuamente: la primera prueba conceptos, la segunda los institucionaliza en servicios permanentes.

Hacia un modelo de Alfabetización en Inteligencia Artificial (AIL)

La síntesis inter-casos permite delinear un modelo AIL de cuatro dimensiones interconectadas, cada una reforzada por puntos de control de equidad para prevenir la brecha de IA:

1. Dimensión técnico-conceptual: Se centra en que el estudiantado comprenda arquitecturas, métricas y limitaciones de modelos —precisión, *recall*, sesgo de entrenamiento—; incluye prácticas de *prompt design* y depuración de datos (Su & Yang, 2023b);
2. Dimensión crítico-interpretativa: Invita a deconstruir el “mito de neutralidad” algorítmica: análisis de sesgos, triangulación de fuentes y evaluación de transparencia, con marcos como la “línea de preguntas críticas” de DeVasto y Palmer (2024);
3. Dimensión aplicada-práctica: Articula herramientas específicas por disciplina: desde generadores de CV (Ponce, 2024) hasta flujos de trabajo híbridos en comunicación organizacional (Pan et al., 2024). Se promueve la co-creación humano-IA bajo criterios de explicabilidad y autoría compartida;

4. Dimensión socio-ético-cívica: Integra la normativa internacional (Sachoulidou, 2024), los derechos digitales y la justicia de datos. Karan y Angadi (2023) argumentan que la escuela, más que adaptar tecnología, debe formar sujetos críticos que interfieran éticamente en su diseño y gobernanza.

Cada dimensión incorpora checkpoints de equidad: diagnósticos de autoeficacia, accesibilidad multiformato, seguimiento de participación de grupos subrepresentados y mecanismos de retroalimentación comunitaria. De esta forma, el modelo se visualiza como una “doble hélice”: mientras las dimensiones técnico-prácticas desarrollan competencias duras, las dimensiones crítico-cívicas garantizan una brújula ética y un ethos inclusivo.

Implicaciones integradas

- Curriculares: integrar módulos AIL obligatorios en los primeros semestres y reforzarlos con aplicaciones disciplinares en cursos avanzados;
- Bibliotecarias: rediseñar servicios de referencia bajo estándares de IA explicable, incluir talleres y laboratorios ciudadanos de IA;
- Regulatorias: alinear políticas universitarias con marcos de riesgo internacionales, desarrollando guías internas de uso responsable;
- Investigación futura: evaluar longitudinalmente el impacto de la doble hélice AIL en la reducción de la brecha de IA y en la generación de capital crítico en comunidades vulnerables.

En adición, los resultados confirman que una alfabetización en Inteligencia Artificial robusta no puede limitarse a impartir habilidades técnicas. Debe articular, en clave de justicia social, la comprensión de los procesos algorítmicos, la capacidad de interrogarlos y la agencia para transformarlos cuando comprometan el bienestar colectivo.

Implicaciones para los actores globales

La adopción de un modelo integral de Alfabetización en Inteligencia Artificial (AIL) trasciende fronteras geopolíticas y niveles de ingreso: la automatización inteligente configura ecosistemas educativos, informativos, regulatorios y productivos en todos los continentes. A partir de la síntesis metaanalítica, se identifican implicaciones estratégicas para cinco grupos de interés cuya actuación coordinada resulta imprescindible para que la alfabetización algorítmica opere como palanca de equidad y no como motivador de nuevas exclusiones.

Sistemas de educación superior

Las universidades de América, Europa, África y Asia comparten el desafío de integrar la IA en planes de estudio heterogéneos sin diluir la dimensión crítica. Los hallazgos de O'Dea et al. (2024) y Weng et al. (2024) demuestran que los currículos más eficaces articulan módulos troncales de AIL con aplicaciones disciplinares situadas; esta lógica puede escalarse globalmente mediante marcos de referencia flexibles que respeten la diversidad cultural y regulatoria. La inclusión de pedagogías críticas continúan siendo la condición de posibilidad para una innovación socialmente responsable, tal como advierten Flierl y Maybee (2020). De manera complementaria, la formación del profesorado requiere itinerarios de desarrollo profesional que combinen microcredenciales técnicas con comunidades de práctica internacionalizadas en las que se compartan casos y rúbricas.

Redes bibliotecarias y de información

La Federación Internacional de Asociaciones de Bibliotecarios e Instituciones (IFLA), al posicionar las bibliotecas como mediadoras del cambio tecnológico, brinda un marco global para alinear servicios, políticas y estándares (IFLA Journal, 2024). Sin embargo, la evidencia de Watkins y Johnson (2024) y de Cox y Mazumdar (2022) confirma que el personal bibliotecario, incluso en países con infraestructuras robustas, todavía carece de competencias sólidas para evaluar y desplegar sistemas de recomendación, discovery o respuesta conversacional basados en IA. Una implicación práctica

es que los consorcios bibliotecarios deberían elaborar kits multilingües de AIL que incluyan guías metodológicas, glosarios, estudios de caso y plantillas de políticas de privacidad. Asimismo, la organización de hackáthones ciudadanos orientados a la co-creación de soluciones algorítmicas transparentes puede atenuar disparidades geográficas de acceso y fomentar la cocreación local de conocimiento.

Organismos reguladores y foros multilaterales

Aunque la Unión Europea ha tomado la delantera con un esquema de estratificación de riesgos algorítmicos (Sachoulidou, 2024), el análisis de Lo (2023) advierte la existencia de un paisaje normativo fragmentado que deja brechas de gobernanza en contextos con infraestructura jurídica incipiente. Para minimizar la carrera hacia el fondo regulatoria, los hallazgos apuntan a la urgencia de articular principios mínimos globales vinculados a transparencia, explicabilidad, derecho a la objeción humana y protección de datos. Los organismos multilaterales dedicados a la educación superior, la ciencia y la cultura pueden servir de puente entre regiones con capacidades legislativas dispares, apoyándose en repositorios de buenas prácticas y en métricas de impacto ético que permitan monitorear la implementación real de la regulación.

Sectores productivos y mercado laboral

Las empresas de base tecnológica y los conglomerados industriales, independientemente de su localización, se benefician de alianzas universidad–industria que ofrezcan prácticas profesionales en entornos de inteligencia híbrida, tal como reflejan los estudios de Coman y Cardon (2024). Estas experiencias impulsan la empleabilidad y aseguran que el estudiantado internalice las dinámicas colaborativas humano–máquina antes de ingresar en un mercado laboral automatizado. A escala global, los clústeres de innovación podrían cofinanciar programas de residencias rotativas en los que estudiantes de diferentes continentes evalúen y adapten soluciones algorítmicas a realidades socioculturales diversas; dicha movilidad favorecería una comprensión intercultural de los sesgos, reforzando la dimensión crítica del modelo AIL.

Sociedad civil, medios y cultura cívica

La brecha de la IA descrita por Wang et al. (2024) y las asimetrías de género detectadas por Zou et al. (2023) evidencian que la alfabetización algorítmica no puede concebirse solo dentro de las aulas. Organizaciones no gubernamentales, grupos comunitarios y medios de comunicación – tradicionales y digitales– desempeñan un papel insustituible en la difusión de narrativas que desenmascaren la opacidad algorítmica, visibilicen afectaciones a derechos y promuevan la rendición de cuentas. Los hallazgos de Porlezza y Schapals (2024) subrayan que la legitimidad democrática se erosiona cuando la automatización de contenidos informativos no se acompaña de mecanismos robustos de verificación. Por ende, la alfabetización algorítmica global exige campañas de comunicación pública que trasladen, en lenguajes accesibles, la relevancia de comprender y cuestionar los sistemas de IA.

Investigación colaborativa y fronteras del conocimiento

Finalmente, la progresión exponencial de la literatura sobre alfabetización de datos mapeada por Yan et al. (2024) indica que las agendas de investigación deben virar hacia estudios comparados de impacto longitudinal que midan la eficacia de los modelos AIL en reducir la brecha de IA a lo largo del tiempo y en contextos socioculturales heterogéneos. Se requieren, además, métricas de equidad que evalúen la participación de grupos históricamente marginados y la distribución de beneficios socioeconómicos derivados de la automatización. La promoción de consorcios de ciencia abierta acelerará la evolución colectiva del campo, en consonancia con los principios de Schüller (2022) sobre alfabetización de datos como derecho cívico.

En conjunto, las implicaciones derivadas del metanálisis confirman que la alfabetización en inteligencia artificial constituye ya un imperativo global cuya viabilidad depende de la co-responsabilidad de instituciones educativas, redes bibliotecarias, entes reguladores, industria y sociedad civil. Sin convergencia de esfuerzos, la IA corre el riesgo de profundizar desigualdades preexistentes; con políticas coordinadas, en cambio, puede operar como catalizador de innovación justa y desarrollo humano sostenible.

Limitaciones del estudio y del modelo de Alfabetización en Inteligencia Artificial

Como todo metaanálisis cualitativo, el presente estudio está sujeto a varias limitaciones que acotan el alcance del modelo de Alfabetización en Inteligencia Artificial (AIL) propuesto y deben ser tenidas en cuenta al interpretarlo.

En primer lugar, existe un sesgo de idioma. Los estudios que superaron los criterios de calidad y pertinencia están publicados en inglés. Esto implica que el corpus se inclina hacia contextos anglófonos o fuertemente influenciados por marcos normativos y académicos del Norte global. La AIL se formula aquí con una vocación latinoamericana, pero la base empírica se nutre sobre todo de experiencias y debates producidos fuera de la región.

En segundo lugar, se reconoce un sesgo disciplinar. La mayoría de los trabajos analizados proceden de campos como la educación, la bibliotecología, la comunicación otros sectores aparecen representados de forma puntual. En consecuencia, las cuatro dimensiones del modelo AIL y la metáfora de la “doble hélice” se encuentran especialmente ajustadas a ecosistemas educativos y de servicios de información, y su transferencia a otros dominios profesionales debe hacerse con cautela y, preferentemente, acompañada de investigaciones complementarias en esos campos.

En tercer lugar, el protocolo de búsqueda privilegia fuentes indexadas en bases de datos internacionales y no agota la literatura gris ni las experiencias locales no publicadas en revistas académicas. Esto puede dejar fuera iniciativas innovadoras de alfabetización en IA desarrolladas en universidades latinoamericanas con menor visibilidad internacional, así como prácticas emergentes documentadas en informes institucionales o proyectos piloto. Aunque esta decisión mejora la trazabilidad y la verificabilidad del corpus, también restringe la diversidad de voces consideradas.

Finalmente, el modelo AIL propuesto debe entenderse como una construcción interpretativa de alcance heurístico y no como una taxonomía

cerrada. Las cuatro dimensiones identificadas y los puntos de control de equidad emergen de la convergencia de patrones en los 29 estudios analizados, pero no han sido todavía validados mediante diseños empíricos específicos (por ejemplo, estudios longitudinales, análisis comparativos entre instituciones o procedimientos Delphi con expertos). Futuras investigaciones deberán someter a prueba la robustez del modelo en contextos universitarios diversos, así como explorar variaciones disciplinares, regionales y culturales que puedan requerir ajustes o extensiones.

Estas limitaciones no invalidan los aportes del estudio, pero sí señalan la necesidad de leer el modelo AIL como una propuesta situada que ofrece un marco de trabajo útil para universidades latinoamericanas, a la vez que invita a ampliar la evidencia empírica y a diversificar las perspectivas implicadas en la conversación sobre alfabetización en inteligencia artificial.

Conclusiones

La evidencia examinada converge en un principio rector: la alfabetización en inteligencia artificial (AIL) eficaz se sustenta en la interdependencia de tres pilares—fluidez de datos, escrutinio ético y pertinencia contextual—cuya articulación trasciende las fronteras nacionales. La acumulación de competencias puramente técnicas sin una conciencia crítica reproduce la “ilusión de neutralidad” algorítmica señalada por Domínguez Figaredo y Stoyanovich (2023); a la inversa, la reflexión ética desvinculada del conocimiento operativo de modelos y métricas conduce a un activismo retórico sin capacidad de incidencia práctica. Esta dialéctica exige un enfoque holístico que las instituciones de educación superior, las redes bibliotecarias y los organismos de políticas públicas deben asumir como responsabilidad compartida a escala mundial.

Hallazgos clave

1. Trayectorias pedagógicas espirales. Desde la educación primaria (Su & Yang, 2023a; Vartiainen et al., 2024) hasta programas profesionales (Wood et al., 2021), la integración temprana y reiterada de la IA

demuestra ser el mecanismo más sólido para consolidar competencias progresivas y evitar la fragmentación curricular;

2. Imperativos socio-éticos globales. Preocupaciones sobre sesgos, transparencia y justicia algorítmica emergen en todos los contextos revisados (Hermann, 2021; Piller, 2024). El AI Act europeo (Sachoulidou, 2024) ofrece una referencia de gobernanza basada en riesgo que se proyecta como estándar de facto para otras regiones;

3. Brecha de la IA. El análisis de Wang et al. (2024) confirma la existencia de segmentos sociales simultáneamente escépticos y subalfabetizados, condición que multiplica su exposición a daños algorítmicos. Estudios de género (Zou et al., 2023) y territorialidad digital ratifican que cualquier agenda de AIL debe anclar puntos de control de equidad;

4. Bibliotecas como catalizadoras. Las bibliotecas universitarias y públicas se perfilan como nodos estratégicos de democratización tecnológica, pero requieren planes sistemáticos de actualización profesional (Cox & Mazumdar, 2022; Watkins & Johnson, 2024).

Implicaciones mundiales

- Educación superior. Las universidades deben adoptar modelos troncales de AIL, reforzados con aplicaciones disciplinares que traduzcan los conceptos abstractos en problemáticas locales, por ejemplo, sesgos en bases de datos de salud en África o automatización agrícola en Sudamérica;
- Redes bibliotecarias. Bajo la guía de la IFLA (2024), los consorcios regionales pueden crear repositorios abiertos de materiales multilingües, favoreciendo la circulación de buenas prácticas entre Norte y Sur global;
- Política pública. En ausencia de normativa homogénea, los países pueden replicar la lógica escalonada del AI Act e incorporar cláusulas

de soberanía de datos para salvaguardar contextos culturales diversos (Lo, 2023);

- Industria y mercado laboral. La creación de pasantías en entornos de inteligencia híbrida, tal como recomiendan Coman y Cardon (2024), facilita la transferencia de conocimientos y fomenta la adaptabilidad profesional en un mercado automatizado transversalmente.

Llamado a la acción global

La alfabetización algorítmica se erige como pilar ineludible de la ciudadanía del siglo XXI. Sin ella, la expansión exponencial de sistemas de IA amenaza con amplificar asimetrías sociales, económicas y epistémicas; con ella, la IA puede transformarse en vector de innovación inclusiva y bienestar colectivo (Schüller, 2022). Para materializar este potencial se requiere:

- Visión sistémica que rompa silos disciplinares y conecte competencias técnicas con marcos de ética pública;
- Financiamiento sostenible que permita a las instituciones más allá del eje global Norte acceder a infraestructuras formativas y recursos actualizados;
- Compromiso cívico que legitime la participación de comunidades locales en la definición de qué IA se diseña, para quién y con qué fines.

En definitiva, participar de manera significativa en una sociedad mediada por IA demanda la convergencia de saberes técnicos y reflexión crítica. Las conclusiones de este metaanálisis proveen un plano conceptual y operativo que puede adaptarse a diversas realidades educativas y culturales. Al movilizar las palancas complementarias de universidades, bibliotecas, industria y marcos regulatorios, la comunidad internacional dispone de la base necesaria para convertir la alfabetización algorítmica en un derecho universal y en un motor de desarrollo humano con justicia y equidad.

Referencias

- Coman, A. W., & Cardon, P. W. (2024). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*. Advance online publication. <https://doi.org/10.1177/23294906241233224>
- Cox, A. M., & Mazumdar, S. (2022). Defining artificial intelligence for librarians. *Journal of Librarianship and Information Science*, 56(2), 330–340. <https://doi.org/10.1177/09610006221142029>
- DeVasto, D., & Palmer, Z. B. (2024). Building critical AI literacy in the business communication classroom. *Business and Professional Communication Quarterly*, 87(4), 557–572. <https://doi.org/10.1177/23294906241253199>
- Domínguez Figaredo, D., & Stoyanovich, J. (2023). Responsible AI literacy: A stakeholder-first approach. *Big Data & Society*, 10(2). <https://doi.org/10.1177/20539517231219958>
- Flierl, M., & Maybee, C. (2020). Refining information literacy practice: Examining the foundations of information literacy theory. *IFLA Journal*, 46(2), 124–132. <https://doi.org/10.1177/0340035219886615>
- Ghodoosi, B., Torrisi-Steele, G., West, T., & Heidari, M. (2024). Perceptions of data literacy and data-literacy education. *Journal of Librarianship and Information Science*. Advance online publication. <https://doi.org/10.1177/09610006241246789>
- Hermann, E. (2021). Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective. *New Media & Society*, 24(5), 1258–1277. <https://doi.org/10.1177/14614448211022702>
- IFLA Journal. (2024). Libraries as catalysts for knowledge, technology, and social resilience. *IFLA Journal*, 50(2), 209–210. <https://doi.org/10.1177/03400352241264683>
- Jarvis, A., Ho, A., & Lim, G. (2024). Impressing artificial intelligence: Automated job-interview training in professional English subjects. *RELC Journal*. Advance online publication. <https://doi.org/10.1177/00336882241245449>
- Jarrahi, M. H., Lutz, C., & Newlands, G. (2022). Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data & Society*, 9(2). <https://doi.org/10.1177/20539517221142824>
- Karan, B., & Angadi, G. R. (2023). Artificial-intelligence integration into school education: A review of Indian and foreign perspectives. *Millennial Asia*. Advance online publication. <https://doi.org/10.1177/09763996231158229>
- Lo, L. S. (2023). AI policies across the globe: Implications and recommendations for libraries. *IFLA Journal*, 49(4), 645–649. <https://doi.org/10.1177/03400352231196172>
- O'Dea, X., Ng, D. T. K., O'Dea, M., & Shkuratskyy, V. (2024). Factors affecting university students' generative AI literacy: Evidence and evaluation in the

- UK and Hong Kong contexts. *Policy Futures in Education*. Advance online publication. <https://doi.org/10.1177/14782103241287401>
- Pan, W., Liu, D., Meng, J., & Liu, H. (2024). Human–AI communication in initial encounters: How AI agency affects trust, liking, and chat-quality evaluation. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/14614448241259149>
- Piller, E. (2024). Inhuman rhetoric: Generative AI and crisis communication. *Journal of Business and Technical Communication*. Advance online publication. <https://doi.org/10.1177/10506519241280594>
- Ponce, T. M. (2024). AI resume writing: How prompt confidence shapes output and AI literacies. *Business and Professional Communication Quarterly*. Advance online publication. <https://doi.org/10.1177/23294906241273317>
- Porlezza, C., & Schapals, A. K. (2024). AI ethics in journalism (studies): An evolving field between research and practice. *Emerging Media*, 2(3), 356–370. <https://doi.org/10.1177/27523543241288818>
- Sachoulidou, A. (2024). Harnessing AI for law enforcement: Solutions and boundaries from the forthcoming AI Act. *New Journal of European Criminal Law*, 15(2), 117–125. <https://doi.org/10.1177/20322844241260114>
- Schüller, K. (2022). Data and AI literacy for everyone. *Statistical Journal of the LAOS*, 38(4), 477–490. <https://doi.org/10.3233/SJI-220941>
- Su, J., & Yang, W. (2023a). Artificial intelligence and robotics for young children: Redeveloping the five big ideas framework. *ECNU Review of Education*, 7(3), 685–698. <https://doi.org/10.1177/20965311231218013>
- Su, J., & Yang, W. (2023b). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 6(3), 355–366. <https://doi.org/10.1177/20965311231168423>
- Van Niekerk, J., Delpont, P. M. J., & Sutherland, I. (2024). Addressing the use of generative AI in academic writing. *Computers and Education: Artificial Intelligence*, 8, 100342. <https://doi.org/10.1016/j.caeai.2024.100342>
- Vartiainen, H., Kahila, J., Tedre, M., López-Pernas, S., & Pope, N. (2024). Enhancing children’s understanding of algorithmic biases in and with text-to-image generative AI. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/14614448241252820>
- Wang, C., Boerman, S. C., Kroon, A. C., Möller, J., & de Vreese, C. H. (2024). The artificial-intelligence divide: Who is the most vulnerable? *New Media & Society*. Advance online publication. <https://doi.org/10.1177/14614448241232345>
- Watkins, T., & Johnson, Q. (2024). AI and machine learning: What to know and how to talk about it to researchers and patrons. *Information Services and Use*. Advance online publication. <https://doi.org/10.1177/18758789241298501>
- Weng, X., Ye, H., Dai, Y., & Ng, O.-L. (2024). Integrating artificial intelligence and computational thinking in educational contexts: A systematic review of

- instructional design and student learning outcomes. *Journal of Educational Computing Research*, 62(6), 1640–1670. <https://doi.org/10.1177/07356331241248686>
- Wood, E. A., Ange, B. L., & Miller, D. D. (2021). Are we ready to integrate artificial-intelligence literacy into medical-school curriculum? Students and faculty survey. *Journal of Medical Education and Curricular Development*, 8. <https://doi.org/10.1177/23821205211024078>
- Yan, C., Wang, H., & Luo, X. (2024). Knowledge mapping of data literacy: A bibliometric study using visual analysis. *Journal of Librarianship and Information Science*. Advance online publication. <https://doi.org/10.1177/09610006241285512>
- Zou, X., Su, P., Li, L., & Fu, P. (2023). AI-generated-content tools and students' critical thinking: Insights from a Chinese university. *IFLA Journal*, 50(2), 228–241. <https://doi.org/10.1177/03400352231214963>

IA generativa para la elaboración de dictámenes en psicología forense: Desafíos éticos, técnicos y jurídicos

Generative AI for the preparation of reports in forensic psychology: Ethical, technical, and legal challenges

Guillermo Ramírez Zavala

guillermormzavala0@gmail.com

*Red Internacional de Estudios de los Problemas Sociales Multidisciplinarios y
Multisectoriales del Análisis de la Criminalidad (REPSMAC)*

ORCID ID 0009-0003-4179-8303

Resumen: La introducción de la inteligencia artificial generativa (IA Gen) en la psicología forense promete agilizar la elaboración de dictámenes periciales y aportar nuevas herramientas de análisis en el sistema jurídico mexicano. Este artículo examina (1) la viabilidad de aplicar IA Gen en dictámenes psicológicos forenses –por ejemplo, como apoyo en la redacción de informes o en el procesamiento de grandes volúmenes de datos–, (2) los riesgos éticos asociados –como sesgos algorítmicos, opacidad, posibles vulneraciones a derechos fundamentales y dilemas bioéticos–, y (3) las implicaciones jurídicas y normativas –incluyendo el marco legal mexicano vigente (Ley General de Ciencia, Tecnología e Innovación 2023) y la necesidad de lineamientos específicos. Se integran hallazgos de literatura científica reciente (2023–2025) y experiencias comparadas en Colombia, España y EE. UU. para contextualizar el debate. A la luz de los principios bioéticos de autonomía, beneficencia, no maleficencia, justicia y conceptos de justicia algorítmica, se realiza un análisis crítico del uso de IA Gen en peritajes forenses. Finalmente, se propone un conjunto de lineamientos técnicos y éticos para un uso responsable de estas tecnologías en la prueba pericial psicológica. Las conclusiones subrayan la importancia de equilibrar innovación tecnológica con salvaguardas éticas y jurídicas,

para aprovechar las ventajas de la IA Gen sin comprometer derechos ni la integridad del proceso judicial.

Palabras clave: Inteligencia artificial generativa; psicología forense; ética; evidencia pericial; sistema jurídico mexicano.

Abstract: The introduction of generative artificial intelligence (GenAI) in forensic psychology holds the potential to streamline the preparation of expert reports and provide new analytical tools within the Mexican legal system. This article examines (1) the feasibility of applying GenAI in forensic psychological assessments—for instance, as support in report writing or processing large volumes of data; (2) the associated ethical risks—such as algorithmic bias, opacity, potential violations of fundamental rights, and bioethical dilemmas; and (3) the legal and regulatory implications—including the current Mexican legal framework (General Law of Science, Technology and Innovation 2023, CNPP) and the need for specific guidelines. Recent scientific literature (2023–2025) and comparative experiences in Colombia, Spain, and the United States are integrated to contextualize the debate. In light of the bioethical principles of autonomy, beneficence, non-maleficence, and justice, as well as concepts of algorithmic justice, a critical analysis of GenAI use in forensic assessments is carried out. Finally, a set of technical and ethical guidelines is proposed for the responsible use of these technologies in psychological expert evidence. The conclusions emphasize the importance of balancing technological innovation with ethical and legal safeguards in order to harness the benefits of GenAI without compromising rights or the integrity of the judicial process.

Keywords: Generative artificial intelligence; forensic psychology; ethics; expert evidence; Mexican legal system.

Introducción

En años recientes la inteligencia artificial generativa (IA Gen), ejemplificada por modelos de lenguaje como *ChatGPT*, ha irrumpido en numerosos campos

profesionales –incluido el ámbito legal– suscitando entusiasmo e inquietud a partes iguales. En el contexto de la psicología forense, que combina saber psicológico y jurídico, la IA Gen podría influir en la elaboración de dictámenes periciales presentados ante tribunales. Un dictamen de psicología forense es la opinión técnica que emite un perito psicólogo, tras evaluar a personas o analizar hechos relevantes en un proceso judicial (p. ej., valorar el estado mental de un acusado, la credibilidad de un testimonio o el daño psicológico en una víctima). Dicha prueba pericial suele ser compleja y determinante en procedimientos penales, familiares o laborales. Por ello, resulta crucial examinar cómo la utilización de IA generativa podría transformar la práctica pericial forense, tanto en sus potenciales beneficios como en sus riesgos.

La viabilidad tecnológica de la IA Gen para asistir en tareas periciales es cada vez mayor, estos modelos pueden *generar texto coherente, resumir documentos extensos, traducir idiomas, identificar patrones* en datos no estructurados, e incluso producir imágenes o voz sintética. En teoría, un sistema de IA Gen entrenado con información psicológica y legal podría ayudar a redactar informes periciales, analizar resultados de tests psicológicos o simular escenarios clínico-forenses. De hecho, ya se emplean algoritmos en la práctica forense: psicólogos y criminólogos utilizan instrumentos actuariales de evaluación de riesgo para estimar probabilidad de reincidencia u otros factores con peso probatorio (Baz, 2023).

La IA Gen representaría un paso adelante al procesar gran cantidad de datos heterogéneos (entrevistas, historial, pruebas psicométricas) y generar un borrador de dictamen o insights útiles para el perito. Un ejemplo ilustrativo es el concepto de *modelos multimodales generativos* capaces de integrar audio, video y texto de evaluaciones psiquiátricas para apoyar diagnósticos e incluso asistir en la producción de informes clínicos (Tortora, 2024). Estas posibilidades anuncian mayor eficiencia y profundidad en los peritajes.

Ahora bien, la adopción de IA Gen en la justicia plantea interrogantes éticas y jurídicas de primer orden. Un dictamen pericial incide en derechos fundamentales (vida, libertad, honra) y debe fundarse en metodología

científica sólida, transparente e imparcial. Surge así la preocupación de si una IA puede generar contenido fiable y exento de sesgos, y cómo garantizar responsabilidad y respeto al debido proceso. Casos recientes alimentan la cautela: en EE. UU., un abogado fue sancionado por presentar un escrito legal elaborado con ChatGPT que contenía citas jurisprudenciales inventadas por la herramienta (Cedillo Lazcano, 2020).

En Colombia, un juez reveló haber usado ChatGPT para redactar partes de una sentencia de tutela en 2023, lo que provocó debate público sobre la legitimidad de tal práctica; la Corte Constitucional colombiana, al revisar el caso en 2024, avaló un uso asistencial de la IA pero estableciendo criterios estrictos (transparencia, verificación, no delegar la función decisoria, etc.) (Gutiérrez, 2023).

En España, la Sala de lo Civil y Penal del TSJ de Navarra examinó en 2024 si un abogado incurrió en mala fe procesal por apoyar su escrito en ChatGPT sin verificar una referencia legislativa errónea; en su auto, el tribunal recalcó que el uso de estas tecnologías exige *especial diligencia y verificación adicional* por parte del profesional, siendo el abogado el último responsable de la exactitud y rigor del documento forense. Estos incidentes evidencian riesgos reales de la IA Gen (p. ej., “alucinaciones” o errores factuales, sesgos en la información) y han motivado respuestas regulatorias iniciales (Barrientos, 2025).

En México, el tema comienza a cobrar relevancia a través un hito jurisprudencial reciente: la tesis aislada II.2o.C.9 K (11a.) del Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito (Queja 212/2025), publicada el 22 de agosto de 2025 en el Semanario Judicial, que por primera vez fija “elementos mínimos” para un uso ético y responsable de IA en sede jurisdiccional con perspectiva de derechos humanos—i) proporcionalidad e inocuidad (restringir la IA a apoyos técnicos, p. ej., cálculos numéricos, sin sustituir el razonamiento jurídico); ii) protección de datos personales; iii) transparencia y explicabilidad (declarar su uso, metodología, datos y trazabilidad para auditoría); y iv) supervisión y decisión humanas—parámetros que, por analogía, resultan transferibles al quehacer pericial: la IA

Gen puede fungir como auxiliar documentado, nunca como “perito” ni como fuente autónoma de juicio experto, y su empleo exige protocolizar insumos, salidas y validaciones, resguardando confidencialidad, cadena de custodia y control de sesgos conforme a los principios de la LGHCTI (ética, inclusión, no discriminación y responsabilidad social).

Esta línea, además, dialoga con referentes internacionales expresamente citados por el propio tribunal (UNESCO 2021; Directrices de IA Fiable de la UE; y el AI Act de la Unión Europea), ofreciendo un marco normativo blando para guiar lineamientos técnico-forenses internos mientras el legislador mexicano desarrolla reglas específicas (El Economista, 2025).

Este artículo aborda de forma integral la viabilidad, los riesgos éticos y las implicaciones jurídicas de la IA generativa en los dictámenes de psicología forense dentro del sistema mexicano. Se estructura en secciones que analizan: (1) el potencial y usos concretos en que la IA Gen podría apoyar al perito psicólogo (viabilidad), (2) los peligros y dilemas éticos que conlleva su empleo sin las debidas precauciones (riesgos éticos), (3) el estado del marco legal aplicable y comparado, incluyendo la situación en México y en otros países que sirven de referencia, (4) una discusión crítica que contrasta beneficios vs. riesgos a la luz de principios bioéticos y de justicia algorítmica, y (5) una propuesta de lineamientos técnicos y éticos para regular el uso responsable de IA Gen en peritajes psicológicos.

Metodología

El presente estudio se basó en una revisión documental y análisis crítico de fuentes especializadas. Se consultaron artículos científicos publicados entre 2019 y 2025 en los campos de psicología forense, inteligencia artificial, derecho y ética, obtenidos de bases de datos académicas (*Frontiers in Psychiatry*, *IDP*, *InDret*, etc.), así como informes de organismos internacionales (UNESCO, ABA) y legislación nacional pertinente. La selección se enfocó en investigaciones sobre IA generativa en contextos forense-legales, estudios empíricos de uso de IA en salud mental y justicia, y ensayos teóricos sobre ética de la IA.

Asimismo, se incluyeron casos jurisprudenciales recientes y pronunciamientos regulatorios de distintos países para un análisis comparado (p. ej., sentencia T-323/2024 de la Corte Constitucional de Colombia, caso del TSJ Navarra 2024 en España, Opinión Formal 512 de la American Bar Association).

Se procedió primero a una lectura exploratoria para mapear los subtemas clave: formas de uso potencial de IA Gen en peritajes, tipos de riesgos identificados (bias, opacidad, etc.), principios ético-jurídicos implicados y situación normativa. Con base en ello, se definió la estructura del artículo y se realizó una lectura analítica profunda de cada fuente relevante, extrayendo evidencias, datos y argumentos significativos.

La información recopilada se organizó temáticamente (viabilidad, riesgos, marco legal, etc.) y se contrastó de forma crítica, identificando convergencias y divergencias entre autores y entre distintas jurisdicciones. Por último, se aplicó un enfoque hermenéutico-jurídico para interpretar los hallazgos a la luz del ordenamiento mexicano vigente y los principios deontológicos de la práctica pericial, formulando propuestas normativas adaptadas a nuestra realidad.

Esta metodología combinó la perspectiva interdisciplinaria (integrando conocimientos de tecnología, psicología y derecho) con un rigor académico en la selección de fuentes (priorizando literatura revisada por pares y normativa oficial). Dicha triangulación garantiza que el análisis y las conclusiones presentadas se sustenten en evidencia actual y confiable. No obstante, dada la rapidez con que evoluciona la IA Gen, se reconoce que este estudio ofrece una fotografía temporal de un fenómeno en desarrollo; se espera que futuras investigaciones sigan enriqueciendo el debate conforme haya nuevas experiencias prácticas y regulaciones sobre la materia.

Resultados

La posibilidad de incorporar IA generativa en el trabajo pericial psicológico es técnicamente factible y presenta varias vías de aplicación. En esencia, la IA

Gen puede funcionar como una *herramienta de apoyo* que potencia las capacidades del perito, más que como un sustituto de su juicio profesional. A continuación, se examinan los usos viables más relevantes:

Asistencia en la redacción de informes periciales

Un dictamen forense suele ser extenso y demandar tiempo en su elaboración. Los modelos de lenguaje (LLMs) podrían ayudar a *esbozar partes del informe*, por ejemplo, redactando antecedentes del caso, explicaciones teóricas o incluso borradores de conclusiones basadas en datos ingresados. Esto podría agilizar la producción de dictámenes, permitiendo al perito concentrarse en el análisis sustantivo.

Estudios recientes señalan que los LLM avanzados pueden generar textos coherentes y estilísticamente adecuados en contextos clínicos, actuando como “copilotos” de escritura científica. En el ámbito psiquiátrico, ya se observó que IA multimodales pueden asistir en la elaboración de reportes clínicos integrando múltiples fuentes de datos del paciente. De manera análoga, un modelo entrenado con informes psicológicos forenses previos podría sugerir redacciones estándar para ciertos apartados (p. ej., descripción de pruebas psicométricas aplicadas, criterios diagnósticos), siempre bajo la revisión final del experto humano (Corvalán, 2021).

Análisis de grandes volúmenes de información

En casos complejos, el perito debe revisar expedientes voluminosos, historiales médicos, declaraciones, etc. La IA Gen combinada con técnicas de procesamiento de lenguaje natural puede *resumir documentos*, extraer menciones clave o detectar inconsistencias en testimonios.

Su capacidad de integrar datos no estructurados (texto libre de entrevistas, notas clínicas, publicaciones científicas) podría ayudar a formar una visión más completa del caso (Diez-Feijóo Varela, 2025). Por ejemplo, un modelo podría sintetizar cientos de páginas de actuaciones judiciales en unos cuantos párrafos resaltando información psicológicamente relevante para la pericia. También podría comparar rápidamente el relato de un

imputado en distintas declaraciones para ver si mantiene consistencia, lo cual es útil en evaluaciones de credibilidad. Esta automatización del “minado” de información permitiría al perito ahorrar tiempo y no pasar por alto detalles importantes.

Evaluación y predicción de riesgos mediante IA

Como se mencionó, los psicólogos forenses ya usan escalas e instrumentos actuariales para estimar riesgo de violencia o reincidencia (Herramientas como HCR-20, VRAG, etc.). La IA Gen, especialmente combinada con *machine learning* discriminativo, puede elevar esa práctica al analizar patrones complejos en bases de datos de delincuentes. En Estados Unidos, algoritmos como COMPAS han sido empleados para predecir riesgo de reincidencia en decisiones pre-sentenciales, si bien con polémicas por sesgos raciales (Muñoz Rodríguez, 2020).

En el entorno forense mexicano, podría entrenarse una IA con datos históricos (p. ej., perfiles psicológicos de agresores condenados, circunstancias del delito) para estimar la probabilidad de conductas futuras o peligrosidad de un evaluado. Este soporte cuantitativo *podría dotar de mayor objetividad* al dictamen en lo referente a riesgo, siempre y cuando se verifique la validez del modelo. Algunos trabajos reportan que los algoritmos predictivos alcanzan performances moderadas (AUC ~0.75) en predicción de reincidencia, útiles como complemento pero no como único criterio. Así, la IA sería viable como sistema de alerta o apoyo estadístico para las conclusiones del perito en materia de riesgo (American Board of Professional Psychology, 2025).

Detección de simulación o malingering

Otra aplicación promisoría es usar IA para identificar patrones sutiles que delaten simulación de trastornos. Investigaciones en psicología han aplicado *machine learning* a medidas comportamentales (como movimientos de ratón o tiempos de respuesta) para distinguir pacientes genuinos de simuladores con alta precisión (hasta 90%+).

Un modelo generativo entrenado con datos de evaluaciones podría, por ejemplo, elaborar hipótesis sobre la congruencia de los resultados de pruebas de personalidad o síntomas reportados, señalando al perito posibles indicadores de fingimiento. Nuevamente, no sería una “prueba” por sí sola, pero sí un *indicador de probabilidad* que oriente al experto a realizar exploraciones más dirigidas (Gordon & Turnbull, 2024).

Simulación de escenarios y formación pericial

Más allá de casos específicos, la IA Gen puede emplearse para *entrenamiento y educación* de peritos. Investigadores han propuesto generar “gemelos digitales” o simulaciones interactivas de pacientes y situaciones forenses, mediante modelos generativos, a fin de que profesionales en formación practiquen sus habilidades diagnósticas y decisorias. Por ejemplo, un entorno virtual podría simular la entrevista con un imputado con trastorno mental, con respuestas generadas por IA, de forma que el psicólogo practique cómo interrogar y detectar inconsistencias éticas. Asimismo, en juicios simulados, la IA podría asumir el rol de testigo o incluso de “juez virtual” para entrenar al perito en la defensa de su dictamen bajo contrainterrogatorio. Estos usos formativos incrementarían la preparación y reducirían errores en la práctica real (Siliceo Portugal, 2024).

La adopción de IA Gen entre profesionales de la salud mental ya es una tendencia emergente. Una encuesta internacional de 2024 encontró que 43% de los profesionales de salud mental encuestados utilizan alguna herramienta de IA, principalmente para labores de investigación y redacción de reportes. Esto indica que casi la mitad ha incorporado IA en tareas como revisar literatura científica, generar borradores o manejar información, lo que sugiere una creciente confianza en estas tecnologías como apoyo diario.

En el sector legal, otra encuesta citada por la ABA muestra un salto en uso de IA en despachos jurídicos de 19% a 79% en solo un año, reflejando la velocidad con que están permeando todas las profesiones vinculadas a la administración de justicia (American Board of Professional Psychology, 2025). Psicólogos forenses en México podrían sumarse a esta ola, aprovechando IA Gen para lidiar con sobrecarga de casos, uniformizar

criterios periciales y reducir la duración de los procesos. Además, dada la escasez de peritos en zonas apartadas, la IA podría *democratizar el acceso* a herramientas de apoyo forense, siempre que haya capacitación adecuada.

En suma, la viabilidad técnica existe y las aplicaciones potenciales son diversas: desde aceleración de tareas burocráticas hasta análisis sofisticados de conducta. La IA Gen bien empleada tiene el potencial de mejorar la calidad y eficiencia de los dictámenes psicológicos forenses, reforzando la base empírica de sus conclusiones. No obstante, cada una de estas promesas viene acompañada de importantes consideraciones éticas y limitaciones que se deben sopesar cuidadosamente antes de incorporar la IA en casos reales, como se detalla en la siguiente sección (Loaiza et al., 2024).

Riesgos éticos y desafíos de la IA generativa en la pericia forense

A pesar de sus ventajas, la IA generativa conlleva riesgos éticos sustanciales en el contexto forense, donde un error puede traducirse en injusticias (condenar a un inocente, revictimizar a una persona, etc.). Es imprescindible identificar y mitigar estos riesgos antes de utilizar IA Gen en la práctica pericial.

Entre los desafíos éticos y prácticos más significativos destacan:

Sesgos algorítmicos y discriminación. Los modelos de IA aprenden de datos históricos que pueden contener prejuicios. En el ámbito penal, esto es crítico: si los datos de entrenamiento reflejan disparidades (p. ej., perfiles de condenados donde cierto grupo étnico está sobrerrepresentado), la IA perpetuará o incluso exacerbará sesgos en sus resultados.

Estudios han demostrado que grandes modelos de lenguaje replican estereotipos de género y raza presentes en sus fuentes: por ejemplo, se ha visto que ChatGPT y GPT-3.5 asocian profesiones prestigiosas con el género masculino y generan con mayor frecuencia contenido violento al referirse a la religión musulmana. En modelos generativos de imágenes, como DALL-

E 2, se halló subrepresentación de mujeres en roles técnicos y sesgos occidentalocéntricos.

Trasladado a psicología forense, un sistema sesgado podría *evaluar como más peligrosa* a una persona solo por pertenecer a un grupo minoritario o atribuir rasgos de violencia basándose en estereotipos raciales, aunque el individuo en cuestión no presente tales indicios de riesgo (Ramírez, 2023).

Esto atentaría contra el principio de igualdad ante la ley y los derechos humanos. De hecho, en EE. UU. ya hubo controversia con el algoritmo COMPAS de riesgo de reincidencia: un análisis reveló que tenía tasa de falsos positivos mayor para personas negras (marcándolos erróneamente como de alto riesgo). Un dictamen pericial generado o apoyado por una IA con sesgo racial podría no solo ser científicamente inválido sino también inconstitucional por discriminatorio.

Por tanto, la presencia de sesgos es uno de los riesgos más graves, ya que minaría la imparcialidad de la justicia e incrementaría la criminalización desproporcionada de colectivos marginados. El perito tiene la obligación ética de asegurar que las herramientas que use no introduzcan discriminación, pero dada la complejidad de los modelos, detectar sesgos ocultos resulta difícil sin auditorías algorítmicas rigurosas (Roa Avella et al., 2022).

Falta de transparencia y explicabilidad (la “caja negra”). Los modelos generativos complejos, basados en redes neuronales profundas, operan como cajas negras cuyo razonamiento interno resulta opaco incluso para sus creadores. En un proceso judicial, esto choca con la exigencia de motivación y explicabilidad de las decisiones (Bejerano, 2024; Rodríguez-Sánchez, 2024). Todo dictamen pericial debe exponer claramente cómo llega a sus conclusiones (metodología, datos considerados, literatura respaldatoria) para que pueda ser controvertido y valorado por el juez. Si un informe se apoya en las salidas de una IA que no son explicables (“porque el algoritmo así lo indicó”), se compromete el derecho de defensa y al debido proceso.

La literatura advierte que muchos modelos actuales no permiten conocer el peso que cada factor tuvo en su output ni las inferencias realizadas.

Esto genera serias dificultades de accountability: ¿quién responde si la IA se equivocó? ¿Cómo refutar un resultado cuyo proceso subyacente no puede desentrañarse? La opacidad, sumada a la posible naturaleza propietaria del software (código cerrado de compañías), agrava el problema (Calderon & Cueto, 2022; Cantos Pardo, 2024).

La ausencia de explicaciones vulnera principios de justicia procedimental y puede volver inadmisibile o poco fiable la prueba. En términos éticos, esto contraviene el deber de transparencia y la obligación del perito de sustentarlo todo en conocimientos comprensibles. Además, la “magia” de la IA puede impresionar indebidamente a jueces o jurados (“efecto aureola tecnológica”), llevándolos a otorgar un peso excesivo a lo que no entienden, lo cual es peligroso. En síntesis, mientras la IA siga siendo en gran medida inexplicable, no debería utilizarse para decisiones que requieren fundamentación clara, como los dictámenes que apoyan sentencias.

Posibilidad de errores factuales y “alucinaciones”. La IA generativa es propensa a inventar datos plausibles pero falsos cuando no encuentra información suficiente, un fenómeno conocido como *alucinación*. Esto se evidenció dramáticamente en el caso del abogado neoyorquino sancionado: ChatGPT generó referencias a fallos judiciales totalmente ficticios pero con apariencia real. En un informe pericial, una IA podría “rellenar” lagunas con contenido inexacto: por ejemplo, citar un estudio psicológico inexistente para respaldar una conclusión, o describir síntomas que el evaluado nunca reportó, si se le pidiera completar un perfil. Si el perito no detecta estas falsedades, el dictamen contendría información espuria que podría inducir a error al tribunal. La verificación de hechos es, pues, crítica. Otra faceta es la generación de *deepfakes*: IA que producen imágenes, videos o audios falsos pero creíbles.

En un contexto pericial, alguien podría presentar (maliciosamente) un audio “deepfake” simulando la voz de un acusado confesando, o una imagen adulterada de la escena del crimen. Si los peritos o jueces confían ciegamente en herramientas de IA para analizar evidencias multimedia, podrían ser engañados. Incluso hay proyectos como *Forensic Sketch AIrtist* que generan

retratos hablados a partir de descripciones, lo cual podría amplificar los sesgos o errores de los testigos originales. Imaginemos un peritaje donde se use tal herramienta: cualquier distorsión en el rostro generado podría llevar a identificar erróneamente a una persona inocente. Por ende, la IA puede introducir falsedades con apariencia de verdad, lo que plantea un riesgo ético de desinformación y requerirá que el perito mantenga una actitud escéptica y comprobatoria frente a las salidas de la IA (Spinak, 2023).

Excesiva dependencia y erosión del juicio profesional. Un peligro más sutil es que, al usar cada vez más la IA como muleta, el perito humano termine reduciendo su participación activa, aceptando acriticamente las sugerencias del modelo. Esta *automatización de la decisión* podría minar la autonomía y pericia del profesional, convirtiéndolo en un mero validador pasivo. Autores como Adela Cortina advierten que no debemos ceder nuestra capacidad crítica y deliberativa a sistemas automatizados. Si un perito delega demasiado en la IA (por ejemplo, permitiendo que redacte casi todo el informe), se corre el riesgo de una “sustitución de la racionalidad humana”, lo cual la Corte colombiana explícitamente señaló que debe evitarse en la administración de justicia (Cortina Orts, 2024).

Además, la interacción constante con modelos generativos podría producir cierta *homogeneización de los informes*: todos con un estilo similar generado por IA, repitiendo los mismos clichés, perdiendo la riqueza del criterio individual y la adaptación a las particularidades del caso. Desde la bioética, esto afecta el principio de autonomía profesional y puede diluir la responsabilidad: el perito podría alegar “eso puso la IA” ante un fallo, intentando eludir culpa. Pero éticamente, el psicólogo forense no puede abdicar de su discernimiento; usar IA exige un equilibrio delicado para que siga siendo el humano quien toma las decisiones finales.

Las guías de la ABA para abogados sobre IA van en esa línea: el profesional debe entender la tecnología y supervisar sus resultados con rigor. Similarmente, los psicólogos deberán resistir la tentación de confiar ciegamente en la IA por comodidad, pues eso puede llevar a errores de juicio, pérdida de pericia con el tiempo (si uno deja de practicar análisis complejo

porque la máquina lo hace) e incluso problemas de ética profesional por negligencia (American Bar Association, 2024).

Problemas de privacidad y confidencialidad de los datos. Los peritajes psicológicos manejan datos personales sensibles (historial psiquiátrico, traumas, conductas delictivas, etc.). Ingresar esta información en una plataforma de IA Gen (muchas de las cuales operan en la nube) podría vulnerar la confidencialidad si no se toman precauciones. Por ejemplo, servicios como ChatGPT almacenan las consultas de los usuarios; cargar detalles de un caso podría implicar que esos datos residan en servidores externos, tal vez ubicados fuera de México, escapando al control del perito y potencialmente accesibles a terceros no autorizados.

Esto transgrede obligaciones de secreto profesional y podría violar la Ley Federal de Protección de Datos Personales. Incluso si la IA opera localmente, existe el riesgo de brechas de seguridad o de que el modelo revele información sensible *aprendida* a partir de casos previos (aunque en teoría los LLM no deberían divulgar datos específicos, ha habido demostraciones de que pueden filtrar partes de sus datos de entrenamiento).

Así, sin protocolos estrictos, el uso de IA puede comprometer la privacidad de evaluados y víctimas. Éticamente, esto contraviene el principio de confidencialidad inherente a la práctica psicológica forense. Además, podría invalidar un dictamen si la parte contraria argumenta que se divulgó indebidamente información o se violó la cadena de custodia de evidencias psicológicas. Por tanto, cualquier implementación tendría que asegurar un entorno seguro para los datos (idealmente sistemas offline o con encriptación robusta) y apego estricto a normas de protección de datos (Mendoza, 2021).

Impacto en la legitimidad y equidad del sistema de justicia. Sumando todos los puntos anteriores, existe un riesgo macro: que un uso imprudente de IA Gen erosione la confianza pública en la justicia. Si los ciudadanos perciben que las decisiones periciales —y por ende judiciales— las toma una máquina incomprensible, o que incorporan sesgos sistémicos, la legitimidad social del sistema penal puede verse afectada. Estudios

criminológicos señalan que la legitimidad influye en el cumplimiento voluntario de la ley y la cooperación ciudadana. Una justicia percibida como “deshumanizada” o “injusta algorítmicamente” podría generar rechazo y menor confianza en peritos y jueces (Moran, 2021).

En España, Baz Cores (2023) reflexiona que la IA debe manejarse con cuidado para no dañar la justicia procedimental, es decir, la impresión de las partes de haber sido tratadas con dignidad, respeto y con decisiones fundamentadas. Si un dictamen forense basado en IA no puede ser cuestionado adecuadamente porque es opaco, o si se sospecha que penaliza más a ciertos grupos por sesgo, las partes podrían sentir indefensión. En última instancia, la introducción de IA Gen sin suficientes garantías podría vulnerar el derecho a un juicio justo (Art. 20 constitucional mexicano) y llevar a más impugnaciones y apelaciones en tribunales superiores, lo que entorpecería el sistema.

En resumen, los riesgos éticos de la IA generativa en peritajes son reales y variados. No se trata de demonizar la tecnología, sino de reconocer sus límites actuales: reproduce prejuicios de su entrenamiento, carece de explicabilidad, puede inventar datos, requiere supervisión constante, y suscita interrogantes sobre responsabilidad y privacidad. Estos desafíos demandan respuestas antes de su aplicación práctica.

La ética profesional y los principios bioéticos son una guía indispensable: *beneficencia* (usar IA solo si trae beneficio claro a la justicia), *no maleficencia* (evitar dañar a individuos con diagnósticos o recomendaciones erróneas), *autonomía* (mantener el control humano y respetar los derechos de las personas evaluadas) y *justicia* (que la IA no introduzca inequidades, sino idealmente ayude a reducirlas). El siguiente apartado profundiza en este marco normativo-ético, incluyendo aportes de autores clave, para fundamentar lineamientos que respondan a tales riesgos (Gonzalez & Martinez, 2020).

Tabla 1

Tipos de uso de IA generativa en peritajes psicológicos y su nivel de aceptabilidad

Categoría de uso	Ejemplos de aplicación	Condiciones mínimas de aceptabilidad
I. Usos aceptables (bajo control humano directo)	- Redacción preliminar de apartados descriptivos del dictamen. - Síntesis y organización de expedientes voluminosos. - Ordenamiento de información clínica, documental o cronológica.	- Revisión exhaustiva por parte del perito. - Declaración explícita del uso de IA en el informe. - Sin procesamiento autónomo de datos sensibles. - No interviene en la interpretación clínica.
II. Usos riesgosos (solo con verificación reforzada)	- Análisis de credibilidad del testimonio. - Evaluación o estimación de riesgo de violencia o reincidencia. - Procesamiento de datos sensibles o confidenciales.	- Validación cruzada con métodos tradicionales. - Documentación de parámetros y outputs. - Verificación independiente de resultados. - No utilizarse como evidencia principal sin sustento adicional.
III. Usos inadecuados o incompatibles con el debido proceso	- Inferir diagnósticos sin evaluación humana. - Determinar imputabilidad, peligrosidad o psicopatología de forma autónoma. - Generar conclusiones periciales automáticas o redactadas íntegramente por IA. - Sustituir la entrevista, la valoración clínica o el contacto directo con las personas evaluadas.	- Prohibido su uso. - Inadmisible en procesos judiciales. - Contraviene principios de autonomía, no maleficencia y debido proceso. - Riesgo elevado de sesgo, opacidad y afectación de derechos.

Nota. Elaboración propia a partir de revisión documental.

Marco jurídico y análisis normativo en México e internacional

En México, el uso de IA generativa en la elaboración de dictámenes periciales aún no cuenta con una regulación específica; por ende, debemos remitirnos al marco general de la prueba pericial, la normativa en ciencia y tecnología, y principios legales aplicables. Destacan los siguientes referentes:

Ley General en materia de Humanidades, Ciencias, Tecnologías e Innovación (LGHCTI, 2023). Esta nueva ley, de alcance

general, incorpora principios rectores para el desarrollo científico-tecnológico en México. Su Artículo 6 enfatiza que el Estado promoverá la ciencia y tecnología velando por la responsabilidad ética, social y ambiental, así como la igualdad, no discriminación, inclusión y pluralidad en esas actividades. También alude al *principio precautorio* ante posibles daños. Si bien no menciona expresamente IA, estos principios son totalmente pertinentes: el desarrollo e implementación de IA (como podría ser entrenar modelos para uso forense) deben evitar causar daño social, discriminación o injusticia. La LGHCTI consagra el derecho humano a gozar de los beneficios de la ciencia sin menoscabo de otros derechos.

Esto sugiere que innovaciones como la IA Gen pueden adoptarse para mejorar procesos (beneficio de la ciencia), pero *no* sacrificando derechos fundamentales en el camino. De hecho, la ley prohíbe la discriminación por motivos como género, origen étnico, condición de salud, etc., al aplicar avances científicos. Una IA sesgada violaría este mandato. La LGHCTI podría servir de base para requerir que cualquier incorporación de IA en el sistema judicial pase por una evaluación ética y cuente con directrices que aseguren su uso responsable.

Además, esta ley faculta a las autoridades para expedir políticas y lineamientos en materia de innovación; un ejemplo podría ser que el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) –creado por esta ley– elabore, junto con el Poder Judicial, guías sobre IA en la administración de justicia. Cabe mencionar que en 2023 se planteó una posible reforma al CNPP para modernizar la prueba pericial, aunque sin enfoque específico en IA; no obstante, a futuro podrían incorporarse previsiones sobre transparencia de herramientas algorítmicas usadas por peritos.

Legislación en protección de datos personales y ciberseguridad.

Aunque tangencial, es relevante. La Ley Federal de Protección de Datos Personales en Posesión de Particulares (LFPDPPP) y su regulación en posesión de sujetos obligados (LGPDPPO) establecen que datos sensibles (como los de salud mental) requieren consentimiento expreso para su

tratamiento, y medidas de seguridad especiales. Un perito que suba información de un evaluado a un servicio de IA en la nube *podría estar vulnerando estas leyes* si no tiene autorización o si el proveedor no garantiza un nivel adecuado de protección.

Además, la Ley General de Archivos y lineamientos del Poder Judicial sobre manejo de expedientes podrían implicar restricciones a sacar información del expediente fuera de los sistemas institucionales. Todo esto sugiere que, legalmente, el uso de IA debe hacerse preservando estrictamente la confidencialidad, quizá mediante convenios específicos con los desarrolladores de IA o usando versiones locales aisladas de internet. De lo contrario, no solo hay un riesgo ético sino legal de sanciones o nulidad de actuaciones por violar la privacidad de la información forense.

Posibles reformas o iniciativas específicas. Hasta la fecha, no existe en México una “Ley de Inteligencia Artificial” o disposición semejante. Sin embargo, en foros académicos y legislativos se ha discutido la necesidad de un marco para IA. Por ejemplo, en 2024 el Senado mexicano organizó foros sobre IA y derecho, y algunos legisladores mencionaron explorar regulaciones inspiradas en el enfoque europeo (donde se clasifica el uso de IA por niveles de riesgo).

Una reforma plausible al mediano plazo podría ser agregar al CNPP (o en una ley de ciencia aplicada a justicia) la obligación de transparencia algorítmica: es decir, si un perito utilizó un algoritmo o IA para llegar a sus conclusiones, que deba revelarlo en su informe y aportar la información técnica necesaria para evaluarlo (parámetros, tasa de error conocida, etc.).

Algo análogo existe en países como EE. UU. bajo estándares tipo *Daubert*, donde el juez actúa como guardián de la fiabilidad científica de las pruebas. En México, los criterios de admisibilidad de la pericia no están tan desarrollados judicialmente, pero bien podría la Suprema Corte emitir criterios en caso de impugnaciones sobre el uso de IA en algún caso futuro. También podría contemplarse la figura de un perito en informática forense/IA que evalúe la corrección de la herramienta utilizada, actuando casi

como “perito de peritos” para dar fe de que la IA opera sin sesgos inaceptables. Estas ideas aún son especulativas, pero apuntan a la dirección de *adaptar el marco jurídico* para brindar certeza.

A nivel internacional, es instructivo revisar cómo otros países y organismos están abordando la IA en la justicia, pues ofrecen modelos a considerar: En Colombia, la Sentencia T-323/2024 de la Corte Constitucional estableció un precedente central al permitir el uso asistencial de IA, pero bajo 12 principios estrictos: transparencia, responsabilidad del usuario, privacidad, no sustitución de la racionalidad humana, verificación, prevención de riesgos, igualdad, control humano permanente, regulación ética, adecuación a buenas prácticas, seguimiento continuo e idoneidad técnica. Además, ordenó crear una guía oficial para la Rama Judicial. Este modelo muestra que es posible innovar sin sacrificar garantías, y aplicado a la pericia implicaría: informar explícitamente el uso de IA en el dictamen, revisar y corregir toda salida generada, proteger datos personales y no delegar nunca la apreciación probatoria a un sistema automatizado.

En España, aunque no existe una ley específica, el Auto 2/2024 del TSJ de Navarra analizó por primera vez el uso indebido de ChatGPT por un abogado y enfatizó que la IA no exime los deberes de diligencia y veracidad, sino que exige verificación adicional. El debate doctrinal, especialmente sobre la “prueba pericial informática” (Velasco, 2023, *Diario La Ley* 95), refuerza que la última palabra en la valoración probatoria debe permanecer en manos del juez. A nivel europeo, el AI Act —en fase final (2023)— clasifica los sistemas de IA usados en justicia como de alto riesgo, imponiendo transparencia, gestión de riesgos, supervisión humana estricta y prohibiciones como el reconocimiento emocional en entornos judiciales. Su entrada en vigor prevista para 2025 obligará a los Estados miembros a adecuar sus prácticas, ofreciendo un marco útil que México podría considerar para evitar rezagos regulatorios.

En Estados Unidos, pese a la ausencia de regulación federal unificada, la American Bar Association emitió la Opinión Formal 512 (2024), que extiende los deberes de competencia, confidencialidad, comunicación

informada y honorarios razonables al uso de IA. Esto implica que el profesional debe entender la tecnología, proteger datos sensibles, explicar riesgos y no atribuirse trabajo generado por IA. En el ámbito judicial, algunos jueces ya exigen declarar si un escrito fue elaborado con IA y presentar verificación de fuentes, como en el caso del juez federal de Texas (2023), para evitar errores como los ocurridos en el caso Avianca. Además, organismos como NIST han desarrollado marcos de gestión de riesgos aplicables al entorno forense, con énfasis en *fairness*, *accountability* y robustez.

Los organismos internacionales también ofrecen estándares relevantes. La UNESCO (2021) propone principios globales como proporcionalidad, seguridad, justicia, inclusión y supervisión humana. La CEPEJ del Consejo de Europa adoptó en 2018 la Carta Ética Europea sobre IA en justicia, que destaca cinco principios: derechos fundamentales, no discriminación, calidad y seguridad, transparencia e imparcialidad, y control humano. Estos instrumentos, aunque no vinculantes para México, trazan una orientación clara: la IA debe ser auditada, trazable, contestable y nunca desplazar la deliberación humana.

Tabla 2

Síntesis comparada de lineamientos, marcos regulatorios y estándares sobre IA en justicia y su relevancia para la pericia psicológica

Jurisdicción / Organismo	Hallazgos y lineamientos clave	Implicaciones para la pericia psicológica
México	Sin regulación específica sobre IA en dictámenes periciales. El uso debe remitirse al marco general de la prueba pericial, la normativa científico-tecnológica y principios constitucionales.	• Todo uso de IA debe cumplir principios constitucionales: debido proceso, dignidad, igualdad y no discriminación.
Marco jurídico general aplicable a IA en peritajes	LGHCTI (2023): Art. 6 establece principios de responsabilidad ética, igualdad, no discriminación, inclusión, pluralidad y principio precautorio. Derecho humano a beneficios de la ciencia sin vulnerar otros derechos. Prohíbe discriminación en la aplicación de tecnología. • Potencial para que CONAHCyT y el Poder Judicial desarrollen guías sobre IA en justicia.	• Prohibición implícita de utilizar IA con sesgos que generen daño o discriminación (mandato de LGHCTI). • Exigencia de consentimiento informado, confidencialidad reforzada

Jurisdicción / Organismo	Hallazgos y lineamientos clave	Implicaciones para la pericia psicológica
	Protección de datos: LFPDPPP y LGPDPPSO exigen consentimiento expreso y medidas reforzadas para datos sensibles. Riesgo legal si un perito sube datos a IA en la nube sin garantías adecuadas. Ley General de Archivos y lineamientos judiciales podrían limitar extracción de información fuera de sistemas institucionales. Reformas futuras: en discusión nacional—posibles normas sobre transparencia algorítmica, criterios tipo Daubert, obligaciones de revelar el uso de IA y parámetros técnicos, y eventual figura de un perito en informática forense/IA para auditar herramientas utilizadas.	y uso de IA en entornos seguros. • Se prevé que la transparencia algorítmica y la trazabilidad sean requisitos futuros para la validez de la prueba pericial. • Necesidad urgente de lineamientos técnicos-éticos para evitar nulidades por violar garantías procesales.
Colombia Corte Constitucional, Sent. T-323/2024	Uso permitido pero altamente regulado. Define 12 principios: transparencia, responsabilidad, privacidad, no sustitución de racionalidad humana, verificación, prevención de riesgos, igualdad, control humano permanente, regulación ética, adecuación a estándares, seguimiento continuo, idoneidad técnica. Ordena elaborar una guía judicial oficial sobre IA.	• Informar en el dictamen si se usó IA. • Verificar y corregir todos los outputs generados. • Proteger datos sensibles bajo estándares reforzados. • No delegar apreciación probatoria a IA. • Modelo útil para guiar lineamientos mexicanos.
España TSJ Navarra, Auto 2/2024	La IA no exime deberes profesionales; exige verificación adicional. Caso pionero sobre uso indebido de ChatGPT por un abogado. Discusión activa sobre fiabilidad de la “prueba pericial informática”. Doctrina enfatiza que el juez debe conservar la última palabra (Velasco, 2023).	• Revisión exhaustiva y diligente de outputs de IA. • Perito sigue siendo responsable técnico integral del dictamen. • Necesidad de mantener control humano decisorio.
Unión Europea AI Act (2023–2025)	Clasifica IA en justicia como alto riesgo. Requiere transparencia, gestión de riesgos, supervisión humana y calidad de datos. Prohíbe puntuación social y limita reconocimiento emocional en ámbitos judiciales.	• La práctica pericial deberá adaptarse a estándares de auditabilidad, explicabilidad y trazabilidad. • Relevante para México como modelo regulatorio prospectivo.
Estados Unidos	Extiende deberes de competencia, confidencialidad, comunicación informada y honorarios razonables al uso de IA.	• Peritos deben acreditar competencia técnica antes de usar IA.

Jurisdicción / Organismo	Hallazgos y lineamientos clave	Implicaciones para la pericia psicológica
ABA, Opinión Formal 512 (2024)	Abogado debe entender tecnología y proteger datos. Jueces federales exigen declarar si un escrito usó IA y proporcionar verificación de fuentes (Texas 2023). NIST desarrolla marcos de riesgo en IA (fairness, accountability, robustez).	• Posible requerimiento de consentimiento informado para procesar datos. • Debe declararse el uso de IA en el dictamen. • Verificación y control documental como estándar de buena práctica pericial.
Organismos internacionales UNESCO (2021); CEPEJ (2018)	UNESCO: principios globales (proporcionalidad, seguridad, justicia, inclusión, supervisión humana). CEPEJ: 5 principios —derechos fundamentales, no discriminación, calidad-seguridad, transparencia-imparcialidad, control humano.	• Reforzamiento del principio de supervisión humana permanente. • Uso de IA debe ser auditado, trazable y contestable. • Estándares éticos útiles como guía base para el contexto mexicano.

Nota. Elaboración propia a partir de revisión documental.

Discusión: ponderando beneficios, riesgos y principios bioéticos

Tras examinar las promesas de la IA generativa y sus peligros, es preciso realizar una valoración crítica: ¿cómo encontrar un equilibrio que permita aprovechar la IA en la psicología forense sin socavar la ética ni la justicia? Esta discusión se centra en dos ejes: (a) la aplicación de principios bioéticos y de justicia algorítmica para evaluar moralmente el uso de IA Gen, y (b) la necesidad de una aproximación prudente, incremental y con controles humanos robustos.

La bioética clásica propone cuatro principios cardinales (Beauchamp & Childress) que aquí ofrecen un marco útil:

Autonomía

Implica respetar la capacidad de decisión de las personas involucradas y la independencia del profesional. En nuestro contexto, la autonomía se concreta en mantener al perito como agente decisor. La IA debe servirle de apoyo,

pero nunca reemplazar su juicio experto ni coartar su libertad de evaluar críticamente cada caso.

Asimismo, desde la perspectiva de las partes (evaluado, acusado, víctima), su autonomía y dignidad exigen que no sean tratados meramente como “datos” para un algoritmo, sino como sujetos cuyo contexto y singularidad se considera. Por ejemplo, un evaluado tiene derecho a que su evaluación psicológica no sea totalmente estandarizada por una máquina, sino que el perito escuche activamente su relato. La autora Adela Cortina subraya que no podemos delegar las decisiones morales a una supuesta racionalidad algorítmica: la deliberación ética y jurídica debe seguir en manos humanas. Solo así se garantiza respeto por la autonomía de la voluntad (p. ej., decidiendo imputabilidad, medidas terapéuticas, etc., caso por caso con sensibilidad) (Suarez-Munoz, 2024).

Beneficencia

Obliga a buscar el bienestar y beneficio, en este caso de la administración de justicia y de las personas involucradas. La IA Gen podría aumentar la beneficencia si acelera procesos (justicia más pronta) y mejora la calidad de las pericias (mayor base empírica, reducción de errores humanos). Una adopción responsable de IA Gen podría, por ejemplo, reducir la carga de trabajo de peritos escasos, permitiendo que los casos se atiendan con menos demoras –lo cual beneficia a víctimas y acusados al acortar la incertidumbre procesal.

También podría ayudar a estandarizar criterios periciales, reduciendo arbitrariedad. Sin embargo, para cumplir genuinamente la beneficencia, la IA debe demostrar que sus aportes efectivamente mejoran la precisión o equidad de las conclusiones. Introducir IA por moda o presión, sin evidencia de su utilidad, sería contrario a este principio. Por ende, habría que validar cada aplicación (quizá mediante estudios piloto controlados) antes de adoptarla formalmente, asegurando que añade valor (p. ej., detecta mentiras más eficazmente que el promedio humano, o identifica riesgo de violencia con mayor acierto) en lugar de entorpecer (García, 2025; Paulo, 2024).

No maleficencia

“Primero, no hacer daño”. Esto es crucial: no maleficencia implica que si la IA generativa conlleva riesgos graves no mitigables, es preferible no usarla. Por ejemplo, si un modelo no puede explicar su razonamiento y podría generar conclusiones sesgadas, entonces introducirlo en un caso penal podría causar un daño (condena injusta o absolución indebida). Hasta que no se minimice ese riesgo con salvaguardas, sería antiético emplearlo. Aquí aplica el principio precautorio de la LGHCTI: ante incertidumbre sobre un posible perjuicio, abstenerse o extremar precauciones.

La no maleficencia exige también *corregir activamente* los sesgos y errores para que la IA no perjudique a poblaciones vulnerables. Por ejemplo, sabiendo que hay sesgo racial en muchos datos criminales, cualquier sistema de riesgo debería calibrarse para eliminar ese sesgo (quizá mediante técnicas de re-muestreo, ajuste de umbrales por grupo, etc.). De lo contrario, su implementación sería maleficente al replicar discriminación. Además, la no maleficencia cubre daños psicológicos: una IA podría etiquetar a alguien erróneamente como psicópata o simulador; tal estigma es dañino. El perito tiene el deber de prevenir esas lesiones morales verificando cuidadosamente las salidas de la IA antes de plasmar nada lesivo en un dictamen (Peñas, 2024).

Justicia

En este contexto, se refiere a la justicia distributiva y procedimental. Supone equidad, trato igualitario y fundamentación consistente. La “justicia algorítmica” extiende este principio: que los algoritmos sean *justos* en sus resultados y *transparentes* en su lógica, para no minar la igualdad ante la ley. Una IA generativa en pericia solo sería aceptable si promueve la equidad o al menos no la socava. Por ejemplo, si ayuda a eliminar el *bias* personal de un perito (todos tenemos prejuicios conscientes o inconscientes), podría incrementar la justicia; pero si añade un nuevo sesgo sistémico, la reduce.

Asimismo, la justicia procedimental implica que las partes puedan entender y confrontar la evidencia; por tanto, un dictamen apoyado en IA debe ser inteligible para no lesionar ese derecho. La transparencia y

explicabilidad son, en el fondo, exigencias de justicia. Como afirma Cortina, una IA debe estar alineada con nuestros valores fundamentales de dignidad y no convertirse en un nuevo mecanismo de dominación acrítica. En suma, desde el prisma de la justicia, la IA Gen debe evaluarse preguntando: ¿Contribuye a decisiones más justas o podría introducir arbitrariedad invisible? Si es lo segundo, sería injusto adoptarla sin remedio (Fernandez, 2019).

Junto a estos principios, se añaden consideraciones de justicia algorítmica: *transparencia, rendición de cuentas y no discriminación*, que ya hemos abordado como ejes éticos pero que conviene reiterar. Una “IA ética” en el campo forense debe ser desarrollada y usada de manera que cumpla con transparencia (auditabilidad de sus datos y procesos), explicabilidad (poder dar razones comprensibles de sus recomendaciones), *accountability* (que siempre haya un responsable humano identificable por sus acciones) y mitigación de sesgos (pruebas y ajustes constantes para evitar resultados discriminatorios). Estos son justamente los principios que instituciones como UNESCO y la UE han puesto sobre la mesa. Incorporarlos al ejercicio pericial es complejo pero necesario para legitimar la IA ante la sociedad (Colomina Saló, Innerarity & Cantero Gamito, 2024).

Enfoque prudente e incremental

A partir de lo anterior, se aboga por un uso gradual y controlado de la IA generativa en psicología forense, nunca abrupto o generalizado sin preparación. Esto significa que, de introducirse, debería empezar en *proyectos piloto* bajo supervisión estricta, en casos de baja complejidad o como experimento paralelo (por ejemplo, el perito hace su dictamen normal y aparte corre la IA para comparar, sin que influya en el fallo, a fin de evaluar su desempeño).

Solo si los resultados muestran concordancia y utilidad, se pasaría a un uso oficial, y aun así limitado a funciones de apoyo bien definidas. En términos prácticos, la IA Gen podría destinarse inicialmente a tareas auxiliares de gabinete (organizar información, redactar borradores no definitivos) pero la evaluación psicológica cara a cara, la interpretación clínica y la emisión de

conclusiones sustantivas deben permanecer en la esfera humana. Al menos mientras la IA carezca de comprensión genuina y empatía, es impensable delegar en ella la apreciación fina de la conducta humana (Gordon, 2025; Ortega, 2025).

La prudencia en la adopción de IA generativa en la justicia se fundamenta en que sus errores pueden ser sutiles pero altamente impactantes, y en que los sistemas judiciales suelen ser conservadores al incorporar evidencia científica novedosa —como ocurrió con el ADN, que requirió largos procesos de estandarización antes de su aceptación rutinaria. En este sentido, conviene actuar con cautela. La comunidad pericial, idealmente a través de colegios profesionales, debería desarrollar códigos de conducta específicos que delimiten qué herramientas de IA son aceptables y bajo qué condiciones. Asimismo, los jueces necesitan formación para comprender las fortalezas y límites de estas tecnologías; sin dicha capacitación interdisciplinaria, valorar un dictamen que incorpora IA puede volverse problemático y generar litigios sobre admisibilidad (López de Mántaras, 2025).

Un enfoque prudente exige también planes de contingencia ante fallas técnicas o sesgos: si se descubre un error en el sistema utilizado por un perito, podrían cuestionarse la prueba o incluso el proceso mismo. Para evitarlo, todo resultado relevante obtenido mediante IA debería verificarse mediante métodos independientes. Así, si un algoritmo sugiere simulación de síntomas, el perito debería corroborarlo mediante pruebas tradicionales de validación o detección de engaño. Esta duplicación metodológica puede ser demandante, pero es necesaria en las primeras etapas de adopción para no depender ciegamente de la tecnología (López Martínez & García Peña, 2024; Vásquez Pérez, 2024).

Las autoridades judiciales, por su parte, tienen el deber de transparentar y justificar cualquier innovación, aclarando que la IA no sustituye a los expertos, sino que los auxilia, y demostrando que su uso incrementa objetividad y reduce errores humanos. La recomendación general es clara: IA generativa, sí, pero con cautela responsable.

La experiencia internacional proporciona referentes valiosos —los principios de Colombia, la ética regulatoria europea y las cautelas estadounidenses— que deben adaptarse al contexto nacional. Un error en su implementación podría afectar gravemente la justicia y la confianza pública, mientras que un acierto permitiría modernizar la labor pericial con mayor rigor y eficacia. Con esta perspectiva, el siguiente apartado desarrolla lineamientos concretos para una adopción responsable de la IA en la prueba pericial (Cuatrecasas, 2024).

Conclusiones

La incorporación de inteligencia artificial generativa (IA Gen) en el ámbito de la psicología forense plantea desafíos sustantivos en términos técnicos, éticos y jurídicos. Frente a este escenario, se propone un conjunto de lineamientos orientados a guiar su utilización de forma responsable, transparente y científicamente fundamentada. Estas directrices, dirigidas a peritos, autoridades judiciales y desarrolladores de tecnología, pueden fungir como base para futuras guías de buenas prácticas institucionales o para la eventual creación de marcos normativos vinculantes. Su estructura se fundamenta en principios de ética profesional, legalidad, debido proceso, y protección de derechos fundamentales, en concordancia con estándares internacionales y nacionales.

Rol asistencial, no decisorio

La IA generativa debe entenderse como herramienta de apoyo técnico y no como sustituto del juicio profesional humano. Toda información procesada por modelos de IA debe ser validada críticamente por el perito, quien conserva la responsabilidad exclusiva sobre las conclusiones del dictamen.

El principio de no delegación de la racionalidad humana, reconocido en decisiones judiciales como las de la Corte Constitucional colombiana, impone que ningún diagnóstico, estimación de riesgo, o inferencia sobre imputabilidad se derive automáticamente de la IA. Su uso puede incluir la elaboración de borradores o el procesamiento preliminar de documentos,

pero toda producción automatizada deberá ser corregida, contrastada y justificada por el experto ante instancias judiciales.

Transparencia y divulgación en el informe pericial

La utilización de IA debe ser expresamente declarada en el informe pericial. Esta declaración debe incluir el propósito del uso, la herramienta empleada, y el grado de intervención del perito en la edición del contenido. Ejemplo de ello sería: “Se utilizó la herramienta X de IA generativa para elaborar un resumen inicial de antecedentes, el cual fue revisado y corregido por el perito”. Adicionalmente, deberá conservarse información técnica relevante (modelo, versión, parámetros) para permitir auditoría y contrainterrogatorio por parte de las partes procesales. La omisión de esta información podría constituir una falta ética y afectar la validez jurídica del dictamen.

Verificación humana y validación cruzada

Todo contenido generado por IA debe ser sometido a un proceso riguroso de verificación humana. Ello implica cotejar datos, identificar errores o inferencias infundadas, y asegurar que las conclusiones sean coherentes con la evidencia empírica y normativa. En casos sensibles, se recomienda la validación cruzada con instrumentos estandarizados. Este procedimiento debe registrarse como parte del expediente técnico del perito, y podrá ser requerido por el juez o las partes como muestra de diligencia profesional.

Capacitación y competencia técnica del perito

El uso de IA requiere habilidades especializadas. En consecuencia, solo los peritos capacitados técnica y éticamente en herramientas de IA generativa deben emplearlas. La capacitación debe abarcar el funcionamiento de los modelos, sus sesgos estructurales, interpretación de outputs probabilísticos, límites jurídicos y salvaguardas éticas. Esta formación es equiparable a la requerida para la aplicación de pruebas psicométricas. La carencia de competencias en este campo podría constituir negligencia profesional en caso de error o afectación al debido proceso.

Selección de herramientas confiables y auditables

Se deben emplear únicamente herramientas que ofrezcan garantías de precisión, protección de datos, capacidad de auditoría y posibilidad de revisión de logs. La validación previa por parte de instituciones judiciales, laboratorios independientes o universidades contribuiría a legitimar su uso. Debe evitarse el uso de modelos opacos o sin certificaciones mínimas de calidad. El criterio de selección deberá documentarse, y estar disponible ante requerimientos judiciales o periciales.

Mitigación activa de sesgos y promoción de la equidad

Antes de su implementación práctica, las herramientas de IA deben ser sometidas a pruebas de sesgo para evitar efectos discriminatorios por género, etnia, edad u otras condiciones. El perito debe asumir una postura activa en la revisión de outputs, identificando estereotipos, lenguaje sesgado o inferencias generalizadas que vulneren el principio de individualización. Las conclusiones no pueden basarse en perfiles poblacionales automatizados, sino en análisis del caso concreto. Este lineamiento se alinea con el principio de igualdad y no discriminación contenido en la Constitución y en tratados internacionales.

Protección de la confidencialidad y seguridad de la información

El tratamiento de datos personales mediante IA requiere un nivel máximo de confidencialidad. Se debe priorizar el uso de sistemas en entornos cerrados y protegidos. En caso de recurrir a plataformas externas, es obligatorio anonimizar los datos y verificar que los términos de uso no impliquen cesión o reentrenamiento con la información ingresada. Los peritos deben actuar conforme a la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), solicitando consentimiento informado o autorización judicial previa. Los archivos generados por IA deben resguardarse con el mismo nivel de seguridad que el expediente clínico o pericial.

Rendición de cuentas y supervisión institucional

La responsabilidad del uso de IA recae exclusivamente en el perito, quien deberá asumir las consecuencias técnicas, éticas y legales de su implementación. Se propone establecer mecanismos de supervisión colegiada o comités de ética judicial para la revisión aleatoria o dirigida de dictámenes asistidos por IA. Asimismo, los peritos deberán conservar registros de las interacciones con las herramientas utilizadas (prompts, outputs, versiones), los cuales podrán ser auditados en instancias procesales. Esto garantiza trazabilidad y posibilidad de apelación fundamentada.

Limitaciones de uso en materias sensibles

Ciertos ámbitos del peritaje forense deben mantenerse, al menos temporalmente, al margen del uso de IA, en atención a su complejidad y sensibilidad. Casos que involucren evaluaciones de imputabilidad penal, infancia, pueblos indígenas o violencia de género, podrían representar contextos donde los riesgos de sesgo, malinterpretación o deshumanización superen los beneficios potenciales de la tecnología. La exclusión podrá definirse por criterios normativos emanados del Consejo de la Judicatura o jurisprudencia especializada.

Actualización y mejora continua de los lineamientos

La naturaleza dinámica de la IA exige que estos lineamientos sean revisados periódicamente. Se sugiere una revisión anual por un comité multidisciplinario que integre expertos en psicología forense, derecho, ética e inteligencia artificial. Este comité incorporará nueva evidencia científica, incidentes registrados y avances tecnológicos. México deberá integrarse activamente a foros internacionales sobre IA y justicia, a fin de armonizar estándares, intercambiar experiencias y anticipar escenarios futuros.

El sistema jurídico mexicano se encuentra en una fase inicial respecto a la regulación de la IA en la justicia, lo que brinda la oportunidad de construir un marco sólido inspirado en experiencias comparadas. Aunque no existe normatividad específica, principios constitucionales como la no

discriminación, el debido proceso y la ética en la ciencia deben orientar cualquier uso de IA generativa en peritajes, mientras que la ética profesional y la prudencia judicial llenan el vacío regulatorio.

En este contexto, se proponen lineamientos técnicos y éticos que preserven la centralidad del juicio humano —el perito como garante de calidad y el juez como filtro de admisibilidad— reforzados por obligaciones de transparencia, verificación, capacitación y salvaguardas contra sesgos y violaciones a la privacidad. La premisa es que la tecnología debe ajustarse a los valores jurídicos, lo cual requiere gobernanza activa y la capacidad de excluir usos que no cumplan estándares de equidad y fiabilidad. Frente a la reticencia inicial hacia estas tecnologías, un marco claro y formación adecuada pueden convertir la incertidumbre en un uso competente y consciente, favoreciendo una integración gradual que aproveche la IA en tareas que realmente aportan valor sin renunciar a la interpretación contextual, la empatía y el razonamiento ético-jurídico.

El objetivo es una justicia aumentada pero humanizada, donde la IA complemente sin sustituir, guiada por principios bioéticos y de justicia algorítmica; como advierte Cortina, “una tecnología sin ética es simplemente poder sin brújula” (2024). Alcanzar este equilibrio exige diálogo interdisciplinario y más investigación sobre sus impactos psicológicos, procesales y sociales, pero los hallazgos permiten afirmar que la IA generativa puede incorporarse sin sacrificar garantías ni principios, siempre que permanezca bajo el control reflexivo de profesionales responsables.

Tabla 3

Recomendaciones diferenciadas para actores clave en el uso de IA generativa en peritajes psicológicos

Actor	Responsabilidades y acciones recomendadas	Justificación ética y jurídica
Peritos psicólogos	• Declarar explícitamente el uso de IA en el dictamen. • Utilizar IA únicamente para tareas auxiliares (borradores, síntesis documental).	• Garantiza transparencia y trazabilidad. • Salvaguarda la autonomía profesional y el debido proceso.

Actor	Responsabilidades y acciones recomendadas	Justificación ética y jurídica
	● Realizar verificación cruzada de toda salida generada por IA. ● Mantener control humano absoluto en la interpretación clínica y en las conclusiones. ● Proteger datos sensibles mediante anonimización y entornos seguros.	● Reduce riesgo de sesgos, errores y afectaciones a derechos fundamentales. ● Cumple con LFPDPPP y principios bioéticos.
Tribunales y autoridades judiciales	● Exigir que todo dictamen indique si se utilizó IA y con qué propósito. ● Solicitar métricas de desempeño o tasa de error cuando la IA aporte elementos analíticos. ● Establecer criterios de admisibilidad y estándares de verificación. ● Capacitar jueces en limitaciones técnicas y éticas de IA generativa. ● Supervisar que la decisión final siempre sea humana.	● Asegura legalidad y debido proceso. ● Facilita el contrainterrogatorio y la defensa técnica adecuada. ● Impide la delegación indebida de funciones jurisdiccionales. ● Previene decisiones basadas en “caja negra” algorítmica.
Colegios profesionales y asociaciones periciales	● Emitir guías de buenas prácticas para uso responsable de IA en pericia psicológica. ● Desarrollar códigos deontológicos específicos para IA generativa. ● Implementar programas de capacitación continua en ética algorítmica. ● Crear comités de revisión técnica de dictámenes asistidos por IA.	● Uniformiza estándares profesionales. ● Minimiza variabilidad arbitraria entre peritos. ● Proporciona marcos de control ético reconocidos por tribunales. ● Promueve la confianza pública en la profesión.
Desarrolladores y proveedores de IA	● Diseñar modelos auditables y compatibles con revisión pericial y judicial. ● Incluir métricas de sesgo, desempeño y explicabilidad del modelo. ● Implementar salvaguardas para evitar uso indebido en contextos forenses de alto riesgo. ● Ofrecer versiones locales o seguras para manejo de datos sensibles.	● Permite trazabilidad algorítmica y auditoría técnica. ● Reduce discriminación algorítmica y riesgo de afectación de derechos humanos. ● Satisface obligaciones legales de protección de datos. ● Facilita la integración segura de IA en el sistema de justicia.

Nota. Elaboración propia a partir de revisión documental.

Referencias

- American Bar Association, Standing Committee on Ethics and Professional Responsibility. (2024). *Formal Opinion 512: Conducting competent and ethical representation in the use of generative AI*. <https://debevoisedatablog.com/wp-content/uploads/sites/2/2024/08/aba-formal-opinion-512.pdf>
- American Board of Professional Psychology. (2025, 24 de marzo). AI, mental health, and forensics: Is this the future? *On Board with Professional Psychology*, 5.
- Baz, O., & Fernández Molina, E. (2022). La justicia procedimental y la legitimidad del sistema penal en España. *European Journal of Criminology*, 19(2), 228–247.
- Baz Cores, O. (2023). Repercusiones de la inteligencia artificial en la legitimidad del sistema penal: Una reflexión desde la justicia procedimental. *InDret*, 4(2023), 291–316.
- Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.
- Bejerano, P. G. (2024, 15 de octubre). La caja negra de la IA que se resiste a los investigadores. *El País*. <https://elpais.com/tecnologia/2024-10-15/la-caja-negra-de-la-ia-que-se-resiste-a-los-investigadores.html>
- Calderón-Ortega, M. A., & Cueto Calderón, C. A. (2022). Prueba por inteligencia artificial: Una propuesta de producción probatoria desde el dictamen pericial científico en Colombia. *Civilizar*, 22(42), e20220106. <https://doi.org/10.22518/jour.ccsch/20220106>
- Cantos Pardo, M. (2024). La algoritmización del dictamen pericial: ¿Puerta de entrada para la aparición del “perito-robot”? *Actualidad Jurídica Iberoamericana*, 21, 434–467.
- Colomina Saló, C., Innerarity, D., & Cantero Gamito, M. (Coords.). (2024, diciembre). *Desigualdad algorítmica: Gobernanza, representación y derechos en la L4* (Núm. 138). CIDOB. <https://www.cidob.org/sites/default/files/2025-02/AFERS%20138.pdf>
- Corte Constitucional de Colombia. (2024, 6 de junio). *Sentencia T-323 de 2024* (M. P. Diana Fajardo Rivera).
- Cortina, A. (2024). *¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnologizada*. Ediciones Paidós.
- Cross, S., Bell, I., Nicholas, J., Valentine, L., Mangelsdorf, S., Baker, S., Titov, N., & Alvarez-Jimenez, M. (2024). Use of AI in mental health care: Community and mental health professionals survey. *JMIR Mental Health*, 11(1), e45140.
- Cuatrecasas Monforte, C. (2022). *La inteligencia artificial en el proceso penal de instrucción español: Posibles beneficios y potenciales riesgos* [Tesis doctoral, Universitat Ramon Llull]. TDX. <http://hdl.handle.net/10803/675100>
- Diario Oficial de la Federación. (2023, 9 de mayo). *Ley General en materia de Humanidades, Ciencias, Tecnologías e Innovación*.

- Fernández Delgado, M. Á. (2019). Interdisciplina y derecho: El ascenso de la tecnarquía y un llamado a deconstruir la Torre de Babel de las profesiones. *Revista de Investigaciones Jurídicas*, 43, 201–236.
- García Garduño, D. J. M. (2025). Entre algoritmos, derechos humanos y principios éticos: El impacto de la inteligencia artificial en la atención médica. *Ecos Sociales*, 13(38), 46–66. <https://doi.org/10.19136/es.v13n38.6609>
- González Arencibia, M., & Martínez Cardero, D. (2020). Dilemas éticos en el escenario de la inteligencia artificial. *Economía y Sociedad*, 25(57), 93–109. <https://doi.org/10.15359/eqs.25-57.5>
- Gordon, S. F., & Turnbull, B. (2024). Adopción de la inteligencia artificial en el campo de la psicología. *Psicología Iberoamericana*, 31(2). <https://doi.org/10.48102/pi.v31i2.547>
- Granda Romero, B. M., Quezada Quezada, J. M., & Durán Ocampo, A. R. (2025). Aplicaciones de la inteligencia artificial en la Criminalística. *Didáctica y Educación*, 16(1), 420–449.
- López de Mántaras, R. (2025, 9 de junio). IA y Justicia: Una combinación que exige prudencia y responsabilidad. *La Vanguardia*.
- López Martínez, F., & García Peña, J. H. (2024). IA y sesgos: Una visión alternativa expresada desde la ética y el derecho. *Revista Iberoamericana de Derecho Informático*, 15(1), 109–121.
- Mendoza Enríquez, O. A. (2021). El derecho de protección de datos personales en los sistemas de inteligencia artificial. *Revista IUS*, 15(48), 179–207. <https://doi.org/10.35487/rius.v15i48.2021.743>
- Miró-Llinares, F. (2023). Digitalización y algoritmización de la justicia: Introducción a un monográfico en tiempos de regulación de la IA. *IDP. Revista de Internet, Derecho y Política*, 39, 1–6.
- Morán Espinosa, A. (2021). Responsabilidad penal de la inteligencia artificial (IA). ¿La próxima frontera? *IUS: Revista del Instituto de Ciencias Jurídicas de Puebla*, 15(48), 289–323.
- Muñoz Rodríguez, A. B. (2020). El impacto de la inteligencia artificial en el proceso penal. *Anuario de la Facultad de Derecho. Universidad de Extremadura*, 36, 695–728. <https://doi.org/10.17398/2695-7728.36.695>
- Ortega Litardo, F. M. (2025). El potencial de la inteligencia artificial en la psicología clínica: Innovaciones que transforman la asistencia psicológica. *Multidisciplinary Journal Educational Regent*, 2(1), 1–12.
- Parmigiani, G., Meynen, G., Mancini, T., & Ferracuti, S. (2024). Editorial: Applications of AI in forensic mental health – Opportunities and challenges. *Frontiers in Psychiatry*, 15, 1153764. <https://doi.org/10.3389/fpsyt.2024.1153764>
- Paulo Neto, A. (2024). Bioética y democracia en la sociedad digital. *Revista de Bioética y Derecho*, 60, 19–34. <https://doi.org/10.1344/rbd2024.60.42990>

- Peñas García, L. (2024, 28 de mayo). La ética en la inteligencia artificial generativa. *Revista Labor Hospitalaria*, 338.
- Ramírez Autrán, R. (2023). Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial: Una revisión documental. *Entretextos*, 15(39). <https://doi.org/10.59057/iberoleon.20075316.202339664>
- Roa Avella, M. del P., Sanabria-Moyano, J. E., & Dinás-Hurtado, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal*, 8(1). <https://doi.org/10.22197/rbdpp.v8i1.615>
- Rodríguez-Sánchez, A. E. (2024). La posibilidad de explicación científica a partir de modelos basados en redes neuronales artificiales. *Revista Colombiana de Filosofía de la Ciencia*, 24(48). <https://doi.org/10.18270/rcfc.4288>
- Starke, G., & Ienca, M. (2023). Out of their minds? Externalist challenges for using AI in forensic psychiatry. *Frontiers in Psychiatry*, 14, Article 1134497. <https://doi.org/10.3389/fpsyt.2023.1134497>
- Suárez-Muñoz, F., & Bautista Tejeda, A. A. (2024). Autonomía de la inteligencia artificial: Una reflexión crítica desde la filosofía kantiana. *SUMMA*, 6(1), 1–11. <https://doi.org/10.47666/summa.6.1.9>
- Tiffon Nonis, B.-N. (2024). *La inteligencia artificial aplicada en la psicología criminal y forense: ¿Reto, realidad o ficción?* Real Academia Europea de Doctores.
- Tortora, L. (2024). Beyond discrimination: Generative AI applications and ethical challenges in forensic psychiatry. *Frontiers in Psychiatry*, 15, Article 1346059. <https://doi.org/10.3389/fpsyt.2024.1346059>
- Tribunal Superior de Justicia de Navarra. (2024, 4 de septiembre). *Auto 2/2024* (Recurso Núm. 17/2024, Sección Penal).
- Vásquez Pérez, M. N. (2024). Ética y género en la IA: Identificar sesgos de género en IA mediante pensamiento complejo. *Revista de Investigación Científica y Crítica Educativa*, 2(2), Artículo 49. <https://doi.org/10.48168/RICCE.v2n2p49>

IA generativa y campañas publicitarias: Retos sociales, éticos y normativos en la transformación algorítmica de la comunicación corporativa

Generative AI and advertising campaigns: Social, ethical, and regulatory challenges in the algorithmic transformation of corporate communication

Fernando Checa García

Fernando.checa@unir.net

Universidad Internacional de La Rioja

ORCID ID 0000-0003-2919-3435

Víctor Núñez Fernández

Victor.nunez@udima.es

Universidad a Distancia de Madrid

ORCID ID 0000-0002-6359-5959

Resumen: La incorporación de sistemas de inteligencia artificial generativa en la comunicación corporativa y las campañas publicitarias está modificando profundamente las dinámicas de producción, distribución y recepción de mensajes persuasivos. Este artículo analiza los impactos sociales, éticos y normativos derivados del uso de IA para generar contenido comercial automatizado, atendiendo a fenómenos como los chatbots promocionales, los influencers virtuales o los anuncios sintéticos. Desde una perspectiva sociológica y crítica, se examina cómo estas tecnologías reconfiguran la confianza en las marcas, alteran las relaciones comunicativas entre empresas y consumidores, y tensionan los marcos regulatorios vigentes en la Unión Europea y en España. Mediante un análisis documental y normativo de carácter cualitativo, se identifican vacíos legales y desafíos de gobernanza algorítmica. Se concluye con propuestas orientadas a garantizar una

comunicación responsable, trazable y socialmente legítima en un entorno mediado por sistemas automatizados de generación de contenido.

Palabras clave: Inteligencia artificial; comunicación social; publicidad; normas jurídicas; ética de la tecnología.

Abstract: The incorporation of generative artificial intelligence systems into corporate communication and advertising campaigns is profoundly transforming the dynamics of production, distribution, and reception of persuasive messages. This article examines the social, ethical, and regulatory impacts arising from the use of AI to generate automated commercial content, focusing on phenomena such as promotional chatbots, virtual influencers, and synthetic advertisements. From a sociological and critical perspective, it explores how these technologies reshape trust in brands, alter communicative relationships between companies and consumers, and challenge the existing regulatory frameworks in the European Union and Spain. Through a qualitative documentary and regulatory analysis, the study identifies legal gaps and challenges related to algorithmic governance. The article concludes by proposing measures aimed at ensuring responsible, traceable, and socially legitimate communication in an environment increasingly mediated by automated content generation systems.

Keywords: Artificial intelligence; social communication; advertising; legal norms; ethics of technology.

Introducción

La expansión de los sistemas de inteligencia artificial generativa (IAG) en la comunicación corporativa y publicitaria está reconfigurando las relaciones sociales en torno a la producción, circulación y recepción de mensajes persuasivos. Herramientas como ChatGPT, DALL·E, Midjourney o

Synthesia permiten generar contenido textual, visual y audiovisual a gran escala, de manera automatizada, en entornos de interacción donde la distinción entre emisores humanos y sintéticos se vuelve cada vez más difusa (Benito García, 2025; Macías, 2025). Esta transformación tecnológica plantea una serie de preguntas centrales, en particular cómo la automatización algorítmica del discurso incide en la legitimidad de los actores emisores, en la percepción de veracidad y autenticidad, y en los marcos normativos que regulan la interacción entre organizaciones y ciudadanía.

El fenómeno debe abordarse no sólo como un proceso técnico o jurídico, sino como un cambio profundo en las estructuras simbólicas que sostienen la acción comunicativa de las organizaciones. La generación automatizada de mensajes persuasivos interpela los conceptos de autoría, intención comunicativa y confianza institucional, desdibujando las fronteras entre lo humano y lo artificial, entre lo real y lo sintético (Fuchs, 2022; Zuboff, 2019). Esta problemática adquiere una especial relevancia en el contexto europeo, donde la regulación de la IA se encuentra en pleno desarrollo, y donde los marcos de gobernanza tecnológica buscan equilibrar la innovación con la preservación de derechos fundamentales y principios democráticos (Peña, 2021; Santana Ramos, 2024).

Este artículo, a través de un análisis comparado de marcos normativos europeos y españoles, se propone examinar de qué manera las transformaciones introducidas por la inteligencia artificial generativa en la comunicación comercial están siendo enfrentadas por los sistemas de regulación contemporáneos. Se analizan los conceptos de transparencia, trazabilidad, consentimiento y responsabilidad como categorías clave para comprender el impacto de estas tecnologías sobre la legitimidad del discurso institucional y corporativo.

Desde el punto de vista metodológico, se adopta un enfoque cualitativo y documental, basado en el análisis de política pública de textos normativos, informes institucionales y literatura científica. La investigación se estructura en torno a tres objetivos principales: en primer lugar, delimitar sociológicamente los efectos de la IA generativa en las campañas publicitarias

y la comunicación institucional; en segundo lugar, evaluar la capacidad de los marcos regulatorios actuales, en especial el AI Act, la Digital Services Act y el RGPD, para abordar estos efectos; y en tercer lugar, proponer una serie de recomendaciones orientadas a fortalecer una gobernanza algorítmica centrada en la transparencia y la justicia comunicativa.

La relevancia de este trabajo radica en su contribución al debate sobre cómo se reconfiguran las estructuras de interacción entre organizaciones y audiencias en un entorno mediado por tecnologías generativas. Frente a enfoques predominantemente jurídicos o técnico-instrumentales, se defiende aquí la necesidad de una mirada crítica que sitúe los discursos generados por IA en el contexto más amplio de las transformaciones del capitalismo digital, la concentración del poder comunicativo y la fragilidad epistémica de las democracias contemporáneas tal como han avanzado Zuboff (2019) y Fuchs (2022).

Marco teórico

Inteligencia artificial generativa como medio comunicativo

La inteligencia artificial generativa (IAG) ha comenzado a configurarse no solo como una herramienta técnica de automatización, sino como un nuevo medio de producción simbólica, capaz de generar contenidos textuales, visuales y audiovisuales que circulan en entornos comunicativos híbridos, donde la frontera entre lo humano y lo algorítmico se vuelve crecientemente difusa (Danesi, 2022). Esta transformación tecnológica redefine las dinámicas de la comunicación institucional y corporativa, al descentralizar la autoría, automatizar la producción discursiva y reformular las relaciones de confianza entre emisor y receptor (Cabrero & Caldevilla Domínguez, 2025; Hinojosa et al., 2024).

La IAG puede entenderse como un vector de mediación que altera las estructuras de legitimidad discursiva, al desplazar la intencionalidad comunicativa desde sujetos identificables hacia sistemas automatizados que reproducen patrones aprendidos. Esta descentralización de la intención y del

control tiene consecuencias sobre la percepción de veracidad, autenticidad y responsabilidad. Como ha sido señalado en la teoría de la mediación tecnológica, los medios no son solo canales, sino actores sociales que participan activamente en la construcción del significado. En este sentido, los sistemas generativos pueden ser entendidos como agentes simbólicos que coproducen cultura, discurso y sentido.

Esta nueva configuración se inserta en un ecosistema digital dominado por las grandes plataformas tecnológicas, donde el poder de configurar la visibilidad, la credibilidad y la emocionalidad de los mensajes no depende exclusivamente de las organizaciones emisoras, sino de las arquitecturas algorítmicas que median su difusión (Poveda & Saltos, 2024). De este modo, la IAG no solo automatiza la producción, sino que amplifica las lógicas de personalización, optimización y rendimiento económico propias del capitalismo de vigilancia.

La atribución de agencia comunicativa a sistemas algorítmicos redefine el contrato simbólico entre marcas y audiencias, desplazando los parámetros de la comunicación institucional hacia nuevas formas de interacción donde la opacidad tecnológica, la automatización de la producción y la manipulación emocional se entrelazan. El resultado es una transformación profunda del espacio público simbólico, en el que la generación automatizada de discursos pone en cuestión los fundamentos de la comunicación ética, responsable y democráticamente gobernada (Fuchs, 2022).

Gobernanza algorítmica, confianza y legitimidad

En el contexto de la comunicación institucional y publicitaria automatizada, la transparencia algorítmica se ha convertido en una condición fundamental para la construcción de legitimidad discursiva. En tanto que los sistemas de inteligencia artificial generativa intervienen en la producción y difusión de mensajes persuasivos, su opacidad técnica y funcional plantea desafíos cruciales para la confianza pública y la responsabilidad comunicativa (Veale & Zuiderveen, 2021). La legitimidad de estos sistemas no depende solo de su

precisión técnica, sino de su capacidad para operar dentro de marcos normativos y simbólicos aceptables socialmente.

Liao y Sundar (2022) han desarrollado el modelo MATCH (Motivation, Ability, Trust, Control and Help) para explicar cómo se construye la confianza en sistemas de IA conversacional. En su propuesta, la confianza se apoya en indicadores percibidos por el usuario que remiten a credibilidad, control, interacción fluida y transparencia explicativa. Esta confianza no es automática, sino que debe ser activamente producida mediante señales discursivas, visuales y funcionales que alineen las expectativas del usuario con el funcionamiento real del sistema. Asimismo, Jacovi et al. (2020) definen la confianza en la IA como una forma de contrato simbólico implícito, donde el usuario delega capacidad de juicio en el sistema bajo la condición de que este actúe conforme a normas previsibles, interpretables y auditables.

En este modelo, los sistemas de IAG no solo generan mensajes, sino que participan activamente en la configuración de subjetividades, preferencias y emociones, al servicio de estrategias de segmentación algorítmica optimizadas por el rendimiento (Piedra Alegría, 2024). La confianza, entonces, no es una categoría individual, sino una construcción política mediada por relaciones de poder entre plataformas, marcas y ciudadanía.

La gobernanza algorítmica aparece como respuesta institucional a estos desafíos. Iniciativas normativas europeas, como el Reglamento de Inteligencia Artificial (AI Act) y la Digital Services Act (DSA), proponen marcos regulatorios para garantizar que los sistemas generativos operen bajo criterios de trazabilidad, explicabilidad y control humano significativo. No obstante, como advierte Veale (2021), existe un riesgo de que estas normas se limiten a gestos simbólicos sin una implementación técnica efectiva, especialmente si no se establecen estándares operativos claros ni capacidades institucionales de auditoría.

En este escenario, la legitimidad de los mensajes generados por IA depende no solo de su contenido, sino del ecosistema sociotécnico que los

sustenta: quién los produce, bajo qué reglas, con qué finalidad y con qué capacidad de intervención pública. La confianza no es un atributo inherente a la tecnología, sino un efecto estructural de la transparencia contextual, de la responsabilidad atribuida y de la capacidad ciudadana para entender y controlar los procesos de mediación algorítmica (Gillespie, 2018; Mittelstadt et al., 2019).

Sociología del conocimiento, tecnología y cultura algorítmica

El desarrollo y despliegue social de la inteligencia artificial generativa no puede comprenderse únicamente desde una perspectiva funcional o técnica, sino que exige una interpretación más amplia de sus implicaciones simbólicas, cognitivas y culturales. Desde la sociología del conocimiento y de la tecnología, diversas corrientes han destacado que la tecnología no es un objeto neutro, sino un artefacto social cargado de valores, que co-evoluciona con las formas de organización institucional, los marcos culturales y las expectativas colectivas. Este principio se resume en el enfoque de mutual shaping o co-constitución entre tecnología y sociedad, según el cual los sistemas técnicos y los contextos sociales se configuran recíprocamente (Bijker et al., 1987; Quan-Haase, 2013).

La IAG constituye, en este sentido, un dispositivo epistemológico y cultural que altera los regímenes tradicionales de producción de conocimiento y de representación simbólica. No se trata simplemente de una herramienta para automatizar tareas, sino de una infraestructura que participa activamente en la generación de sentido, modificando las condiciones de posibilidad de la comunicación, la deliberación y la imaginación colectiva.

Una contribución clave en esta línea es el concepto de cultura algorítmica formulado por Ted Striphas (2015), quien sostiene que los algoritmos no solo seleccionan información, sino que modelan activamente los hábitos de consumo cultural, la percepción de relevancia y las formas de atención. Bajo este marco, los contenidos generados por IA no son simples outputs funcionales, sino expresiones de una lógica cultural codificada, que influye sobre los criterios de valor, originalidad y pertinencia. La cultura

algorítmica redefine así las gramáticas simbólicas dominantes, desplazando progresivamente las formas tradicionales de autoridad textual, experiencia estética y juicio comunicativo.

Sociología de la regulación tecnológica

La regulación de la inteligencia artificial en Europa, particularmente en lo relativo a la IA generativa aplicada a la comunicación institucional y comercial, debe ser analizada no solo como una respuesta jurídica a un problema técnico, sino como un instrumento de reorganización del poder simbólico y comunicativo. Desde la perspectiva de la acción reguladora, la normativa actúa como un dispositivo que estructura expectativas, define marcos de legitimidad y redistribuye autoridad entre actores sociales, económicos y políticos (Yeung, 2018).

La Digital Services Act (DSA) y el Reglamento Europeo de Inteligencia Artificial (AI Act) se presentan como pilares de una gobernanza algorítmica orientada a reducir la opacidad de los sistemas automatizados, estableciendo obligaciones de etiquetado de contenido generado por IA, trazabilidad de procesos de personalización, y derechos de explicación y supervisión por parte de los usuarios (Domínguez, 2023; Rodríguez, 2024). Estos textos normativos incorporan explícitamente preocupaciones relativas a la equidad informativa, la veracidad del discurso automatizado y la protección frente a manipulaciones comerciales encubiertas.

No obstante, la literatura ha subrayado que este tipo de regulaciones no puede entenderse únicamente desde su dimensión prescriptiva. Como insisten Veale y Zuiderveen (2021) o Hildebrandt (2022), la regulación tecnológica se inscribe en regímenes sociotécnicos de poder, donde normas jurídicas, infraestructuras técnicas y prácticas institucionales co-producen formas de control y gobierno algorítmico. Así, lo que está en juego no es solo el cumplimiento legal, sino la definición de lo que cuenta como legítimo, seguro o explicable en la comunicación pública automatizada.

Desde la teoría del policy feedback, desarrollada en la sociología política para explicar cómo las políticas públicas configuran identidades, prácticas y expectativas, puede afirmarse que la regulación de la IA está generando un nuevo campo discursivo donde los actores institucionales, corporativos y ciudadanos renegocian sus roles y responsabilidades (Borrás & Edler, 2020; Pierson, 1993). En el ámbito específico de la comunicación digital, esto implica una reconfiguración de las formas de confianza institucional, de las atribuciones de responsabilidad y de los criterios de verdad en los mensajes distribuidos por marcas e instituciones públicas.

Metodología

Este estudio se inscribe en una perspectiva cualitativa de tipo interpretativo, ubicada dentro de los estudios sociales de la tecnología y orientada al análisis crítico de los marcos regulatorios que configuran el uso de la inteligencia artificial generativa. Se adopta aquí un enfoque metodológico que considera la norma como artefacto sociotécnico, es decir, como producto institucional que refleja, articula y reproduce determinadas lógicas culturales, estructuras de poder y modelos de gobernanza (Jasanoff, 2004).

La estrategia de investigación se basa en lo que puede denominarse un análisis de políticas, una metodología cualitativa orientada a comprender los significados, tensiones y efectos que emergen en torno a la regulación de tecnologías disruptivas. Esta perspectiva ha sido utilizada en investigaciones previas centradas en la gobernanza de datos, algoritmos y plataformas digitales, y permite observar cómo las normas jurídicas no solo prescriben comportamientos, sino que también reconfiguran la legitimidad de los actores, la acción institucional y los límites de lo aceptable en la esfera pública digital (Hildebrandt, 2022; Yeung, 2018).

Para el desarrollo del trabajo se han incluido fuentes primarias y documentos institucionales, seleccionados por su relevancia directa para el ámbito de la comunicación publicitaria/corporativa con IAG así como su vigencia en la fecha de cierre del estudio, noviembre de 2025. Se excluyeron proyectos no consolidados, borradores sin versión final y documentos

sectoriales sin validez general. Las versiones citadas corresponden a textos consolidados con URL estables.

Respecto al procedimiento analítico, se ha aplicado un análisis documental cualitativo con una matriz de categorías predefinidas (transparencia, trazabilidad, equidad comunicativa) y una capa transversal de “autoría/etiquetado de contenido sintético” y “explicabilidad”. Se ha realizado una triangulación entre norma, documentos de autoridad y literatura académica para identificar vacíos operativos y ambigüedades.

Atendiendo a las limitaciones de la investigación cabe destacar que el estudio desarrollado es documental, priorizando el ámbito UE/España y no evalúa cumplimiento empírico. Las inferencias sobre campañas se formulan como implicaciones regulativas plausibles; futuras fases incorporarán auditorías de piezas, análisis de etiquetado y entrevistas con responsables de cumplimiento normativo.

La pregunta de investigación que estructura este trabajo es la siguiente: ¿cómo están respondiendo los marcos normativos europeos y españoles a los impactos sociales, simbólicos y organizativos derivados del uso de inteligencia artificial generativa en la comunicación estratégica? A partir de ella, se han formulado tres objetivos específicos de carácter analítico:

O1. Caracterizar el fenómeno de la automatización discursiva en el ámbito publicitario y corporativo desde un enfoque sociológico de la comunicación.

O2. Analizar los marcos normativos e institucionales europeos y españoles como configuradores del entorno de gobernanza algorítmica en este campo.

O3. Evaluar en qué medida estos marcos responden a los principios de transparencia, trazabilidad y equidad comunicativa exigidos por una ética pública digital.

La investigación se ha desarrollado a partir de un diseño documental articulado en torno a un corpus normativo integrado por los principales textos legales de ámbito europeo y español que inciden sobre el uso de sistemas de IA en entornos comunicativos. A nivel europeo, se han seleccionado el Reglamento General de Protección de Datos (RGPD), la Digital Services Act (DSA) y el Reglamento de Inteligencia Artificial (AI Act). Del ámbito español, se han incluido la Ley General de Publicidad (Ley 34/1988), la Ley de Servicios de la Sociedad de la Información (LSSI), la Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales (LOPDGDD) y la Estrategia Nacional de Inteligencia Artificial (ENIA, 2020). Esta selección responde a su relevancia jurídica directa y a su capacidad estructurante de las prácticas comunicativas institucionales. De igual manera se ha analizado el corpus institucional, compuesto por informes de organismos reguladores y autoridades competentes (Comisión Europea, AEPD, CNMC, Parlamento Europeo).

En primer lugar, se ha efectuado una lectura interpretativa de los textos normativos y técnicos, orientada a identificar los marcos que estructuran la representación institucional del problema de la IA generativa en comunicación: ¿qué riesgos se visibilizan?, ¿qué valores se jerarquizan?, ¿qué soluciones se proponen?, ¿quiénes son los actores legítimos? En segundo lugar, se ha procedido a una evaluación cruzada entre los objetivos normativos explícitos y los hallazgos teóricos y empíricos de la literatura especializada, con el fin de detectar vacíos regulatorios, ambigüedades normativas y posibles disonancias entre regulación y práctica institucional (Borrás & Edler, 2020; Montero, 2018). Esta metodología se alinea con la tradición de la sociología de la acción reguladora, que entiende las normas no como objetos estáticos, sino como productos de procesos sociales e institucionales que condensan disputas, intereses y valores.

El estudio ha evitado la incorporación de trabajo de campo empírico directo en esta primera fase, ya que el objetivo principal ha sido establecer una base teórica, normativa y metodológica robusta sobre la cual se puedan construir futuras investigaciones empíricas. No obstante, la investigación incluye una sección de discusión donde se derivan propuestas políticas y

líneas de intervención a partir del análisis realizado, abriendo la posibilidad de implementar en el futuro metodologías mixtas, entrevistas cualitativas o análisis comparativos entre modelos regulatorios.

Resultados

Automatización discursiva y configuración simbólica en el marco normativo

El análisis del Reglamento General de Protección de Datos (RGPD) permite observar que, aunque su origen se sitúa antes del auge de la IA generativa, ya anticipaba la existencia de procesos automatizados de toma de decisiones que afectaban significativamente a las personas (art. 22). El texto adopta una lógica esencialmente procedimental y protectora de derechos individuales, centrada en la protección de datos personales y en el consentimiento informado. El art. 22 RGPD se refiere a decisiones exclusivamente automatizadas con efectos jurídicos o similares; su aplicabilidad a segmentación publicitaria requiere evaluación contextual del impacto y del grado de intervención humana. Desde una perspectiva sociológica, el RGPD sitúa el foco sobre la relación individuo-plataforma, dejando en segundo plano la dimensión estructural y discursiva de la automatización comunicativa.

En cambio, el Reglamento Europeo de Inteligencia Artificial (AI Act) introduce un cambio relevante: clasifica los sistemas de IA según su nivel de riesgo y define a la IA generativa como una tecnología de “riesgo limitado”, a la que se aplican obligaciones específicas de transparencia y divulgación. Esta clasificación construye una narrativa institucional en la que la automatización del contenido no es, per se, problemática, siempre que sea comprensible y etiquetable. La IA generativa se configura así como un actor neutral, cuyo impacto depende de su grado de visibilidad. Esta concepción implica una cierta banalización de los efectos comunicativos estructurales de la automatización, reduciendo el problema a su dimensión informativa. Sin embargo, el AI Act impone obligaciones de transparencia, incluido el etiquetado de contenido sintético o “deepfakes”, y requisitos adicionales

como documentación sobre datos de entrenamiento con derechos de autor y mitigaciones de riesgos.

Por su parte, la Digital Services Act (DSA) centra su regulación en los grandes intermediarios digitales. En el ámbito publicitario, introduce importantes requisitos de transparencia sobre el origen del contenido, la identificación del anunciante y los parámetros de segmentación utilizados. Aunque no menciona explícitamente la IA generativa, su lógica reguladora apunta hacia una visibilización de los dispositivos técnicos que median el discurso, lo que implica un reconocimiento de la automatización como actor relevante en la configuración de las campañas. Sin embargo, desde el punto de vista de la legitimidad, la DSA delega buena parte de la responsabilidad en las propias plataformas, lo cual refuerza la asimetría de poder entre productores institucionales de contenido y entidades privadas. Debe subrayarse que la DSA no exige un etiquetado general de “contenido generado por IA”. Sus obligaciones de transparencia se concentran en publicidad (anunciante, parámetros de segmentación) y en sistemas de recomendación de plataformas muy grandes.

En España, la Ley General de Publicidad (1988) y la Ley de Servicios de la Sociedad de la Información (LSSI, 2002) ofrecen una imagen desfasada del ecosistema comunicativo actual. Aunque ambas incluyen disposiciones relevantes sobre la identificación de la publicidad, el consentimiento y la protección frente a la manipulación, no contemplan la posibilidad de una producción discursiva no humana, ni los mecanismos algorítmicos de personalización automatizada. Esta omisión refleja una inercia normativa, que continúa concibiendo la comunicación institucional desde un paradigma analógico.

En cambio, la Estrategia Nacional de Inteligencia Artificial (ENIA, 2020), aunque no es jurídicamente vinculante, incorpora un lenguaje más actual, que reconoce la capacidad de la IA para intervenir en procesos comunicativos. No obstante, su carácter programático y su falta de especificidad sobre la automatización publicitaria limitan su capacidad para actuar como herramienta estructurante. A ello se suma la ausencia de

referencias al plano simbólico o cultural de la comunicación automatizada, lo cual refuerza una visión instrumental y tecnocentrista.

Como puede observarse en la Tabla 1, los marcos normativos presentan una representación limitada y fragmentada del fenómeno de la automatización discursiva. Mientras las normas europeas más recientes comienzan a construir un lenguaje regulador que reconoce la existencia de sistemas generativos, las leyes estatales españolas mantienen una visión desactualizada del ecosistema mediático. Esta desalineación entre práctica tecnológica y norma legal genera un vacío institucional, que dificulta la construcción de confianza, responsabilidad y legitimidad en el uso de sistemas generativos para fines publicitarios o corporativos.

Tabla 1

Representación del fenómeno de automatización discursiva en los textos normativos analizados

Norma analizada	Reconocimiento de la IA generativa	Enfoque regulador	Dimensión simbólica considerada	Actualización discursiva
Reglamento General de Protección de Datos (RGPD)	Implícito (decisiones automatizadas)	Protección de datos y consentimiento	Baja	Parcial (anterior a la IAG)
Reglamento de Inteligencia Artificial (AI Act)	Explícito (riesgo limitado)	Transparencia y clasificación de riesgos	Media	Alta (mención específica a IAG)
Digital Services Act (DSA)	Implícito	Transparencia en contenidos y segmentación	Media	Alta (plataformas y publicidad)
Ley General de Publicidad (España)	Ausente	Publicidad tradicional e identificabilidad	Nula	Muy baja
LSSI (España)	Ausente	Servicios digitales y consentimiento comercial	Nula	Muy baja
LOPDGDD (España)	Implícito (automatización y datos)	Protección de derechos digitales	Baja	Media
ENIA (España)	Implícito (reconocimiento genérico)	Fomento ético de la IA	Media	Alta (pero no vinculante)

Nota. Elaboración propia.

Marcos normativos como configuradores de la gobernanza algorítmica en comunicación institucional

En el ámbito de la Unión Europea, el RGPD funciona como la base estructural de esta gobernanza, en tanto establece los principios que rigen el tratamiento automatizado de datos personales. Aunque el RGPD no fue diseñado para abordar directamente la generación de contenidos mediante IA, su artículo 22 limita el uso de decisiones automatizadas sin intervención humana, lo cual puede interpretarse como una restricción indirecta a ciertas formas de automatización comunicativa, especialmente en contextos de segmentación emocional o política. No obstante, su efectividad en este campo depende de la interpretación que hagan los reguladores y tribunales, y de la capacidad de los individuos para ejercer derechos que, en la práctica, resultan opacos o difíciles de implementar.

En contraste, el AI Act representa un intento explícito de instaurar una gobernanza algorítmica europea basada en un modelo de regulación por niveles de riesgo. Este enfoque dota al sistema legal de cierta flexibilidad, pero al mismo tiempo reproduce una lógica tecnocrática, en la que los riesgos se miden con criterios esencialmente funcionales y no simbólicos. La IA generativa se ubica en el nivel de “riesgo limitado”, lo que implica obligaciones de transparencia pero no de control institucional fuerte. Esta clasificación tiene implicaciones relevantes: al definir qué tecnologías requieren intervención pública y cuáles no, el AI Act institucionaliza una jerarquía de vigilancia regulatoria, donde la comunicación automatizada queda en un segundo plano.

La Digital Services Act (DSA) introduce una capa complementaria de gobernanza centrada en la supervisión de las grandes plataformas digitales. Su lógica no se basa en la tecnología, sino en la posición estructural del actor: cuanto mayor poder de intermediación tiene una plataforma, más obligaciones le impone la norma. En este sentido, la DSA reconoce que la gobernanza algorítmica no puede depender exclusivamente del tipo de herramienta utilizada, sino también de la capacidad de amplificación

simbólica de los actores que la emplean. Sin embargo, su modelo de autorregulación supervisada y los mecanismos voluntarios de cumplimiento refuerzan el papel de las plataformas como co-gobernantes del espacio público digital, con escaso control externo efectivo (Aminahuel & Rodríguez, 2024).

Nuevamente, en España, los marcos normativos muestran una menor capacidad configuradora. La LOPDGDD reafirma los principios del RGPD y refuerza ciertos derechos en el entorno digital, como el derecho a la portabilidad y la limitación del tratamiento de datos personales. Sin embargo, no establece un marco específico de gobernanza para la automatización comunicativa, lo que deja en manos del mercado o de las plataformas la decisión sobre los estándares aplicables.

Por su parte, la LSSI y la Ley General de Publicidad continúan articulando un modelo de gobernanza fragmentada y reactiva. En lugar de anticipar los desafíos de la automatización simbólica, estas normas responden a violaciones concretas (publicidad engañosa, spam, etc.) sin ofrecer criterios claros sobre la arquitectura institucional deseable. Desde una perspectiva de policy feedback (Pierson, 1993), puede afirmarse que estas leyes generan un ciclo débil de retroalimentación normativa, en el que la regulación no transforma significativamente las prácticas que pretende supervisar.

La Estrategia Nacional de Inteligencia Artificial (ENIA), sin carácter normativo, contribuye a institucionalizar una visión de gobernanza algorítmica centrada en la innovación, la ética y la competitividad. No obstante, el texto omite definir mecanismos específicos de auditoría, rendición de cuentas o protección simbólica frente a la automatización discursiva. Ello genera una narrativa en la que la confianza en la IA se convierte en un objetivo deseable, pero se desvincula de las estructuras institucionales que podrían garantizarla efectivamente.

En la tabla 2 puede observarse cómo el panorama revela que la gobernanza algorítmica en el ámbito comunicativo está marcada por una tensión entre centralización reguladora y delegación operativa. Mientras que

los textos europeos comienzan a establecer un marco común, su implementación se deja en manos de actores privados, cuya lógica responde a intereses comerciales más que a criterios de justicia comunicativa. A su vez, la regulación española carece de herramientas estructurantes, lo que agrava la dependencia del mercado.

Tabla 2

Capacidad configuradora de gobernanza algorítmica según los marcos normativos analizados

Norma analizada	Modelo de gobernanza adoptado	Ámbito de aplicación directa a IAG	Grado de intervención pública	Papel de actores privados	Evaluación sociológica
RGPD	Protección de derechos individuales	Parcial (automatización de decisiones)	Medio	Limitado (consentimiento y reclamación)	Enfocado en datos, no en discurso; gobernanza legal, no simbólica
AI Act	Clasificación de riesgo y transparencia	Directo	Bajo–medio	Alto (cumplimiento y autocertificación)	Prioriza función sobre impacto simbólico; enfoque tecnocrático
DSA	Supervisión estructural de plataformas	Indirecto	Medio–alto	Muy alto (autorregulación supervisada)	Reconfigura relaciones de poder comunicativo; falta control externo
LOPDGDD	Refuerzo del RGPD	Indirecto	Medio	Bajo (extensión de derechos, no estructura nueva)	Gobernanza fragmentaria; sin lógica sistémica
LSSI	Reactiva y sancionadora	Nulo	Bajo	Bajo (sólo sujetos sancionables)	Desfase normativo frente a nuevas prácticas discursivas
Ley General de Publicidad	Supervisión formal de mensajes comerciales	Nulo	Bajo	Bajo (marco tradicional)	Modelo analógico; ignora automatización simbólica

Norma analizada	Modelo de gobernanza adoptado	Ámbito de aplicación directa a IAG	Grado de intervención pública	Papel de actores privados	Evaluación sociológica
ENIA	Visión estratégica de la IA	Genérico	Muy bajo (no vinculante)	Alto (fomento ético-autogestionado)	Retórica ética sin mecanismos operativos

Nota. Elaboración propia.

Evaluación normativa de los principios de transparencia, trazabilidad y equidad comunicativa

El tercer objetivo analítico de este estudio consiste en evaluar en qué medida los principales marcos normativos e institucionales vigentes contribuyen a garantizar tres pilares fundamentales para una comunicación automatizada legítima: la transparencia en la producción de mensajes generados por IA, la trazabilidad de los procesos que intervienen en su circulación y personalización, y la equidad comunicativa entendida como derecho colectivo a una distribución simbólica no discriminatoria ni manipulativa.

En primer lugar, el análisis de la transparencia revela una profunda asimetría entre los textos normativos. El AI Act es el único que establece de forma expresa la obligación de informar a los usuarios cuando están interactuando con un sistema generativo de IA, así como de etiquetar los contenidos como "generados automáticamente" cuando se difunden en entornos públicos. A pesar de esto, dicha obligación se presenta de forma genérica y sin especificar los estándares semióticos, tecnológicos o lingüísticos necesarios para que el etiquetado sea efectivo. Ello apunta a una transparencia formalizada pero no comunicativamente significativa, lo que socava su impacto sobre la percepción y comprensión del usuario (Mittelstadt et al., 2019).

La DSA, por su parte, incorpora obligaciones de transparencia en el ámbito de la publicidad digital, como la identificación del anunciante y la explicación de los criterios de segmentación algorítmica utilizados. Aunque estas disposiciones son relevantes para desentrañar la lógica económica de la comunicación automatizada, no abordan la dimensión semiótica de los

contenidos generados por IA. Tampoco exigen etiquetado explícito cuando la pieza comunicativa es sintética. En este sentido, la DSA prioriza una transparencia funcional, centrada en los procesos de difusión, pero sin atender al carácter algorítmico de la autoría del mensaje.

El RGPD incorpora principios generales de transparencia, especialmente en lo referido al tratamiento de datos y a las decisiones automatizadas. Sin embargo, se focaliza en la trazabilidad de la información personal, y no del contenido en sí. Además, su estructura centrada en el sujeto individual limita su eficacia para garantizar transparencia relacional, es decir, aquella que permite a los colectivos interpretar críticamente la procedencia y objetivos de los discursos automatizados (Pérez Ugena, 2024).

En cuanto a la trazabilidad, el marco regulatorio europeo carece aún de mecanismos estructurados que permitan reconstruir el itinerario de un mensaje generado por IA desde su origen técnico hasta su difusión pública. A pesar de que tanto el AI Act como la DSA apelan a la necesidad de explicabilidad, esta se interpreta desde una óptica técnica o legalista, sin considerar su dimensión comunicativa. La ausencia de registros públicos, estándares interoperables y sistemas de auditoría accesibles dificulta la trazabilidad real de los contenidos automatizados, especialmente cuando estos se integran en entornos programáticos u omnicanales.

En el caso español, la LOPDGDD hereda esta orientación hacia los derechos informacionales individuales, sin incorporar dispositivos específicos para rastrear la circulación simbólica de los mensajes generativos. La LSSI y la Ley General de Publicidad carecen completamente de herramientas normativas para garantizar la trazabilidad de los mensajes producidos por IA. Este vacío normativo configura una infraestructura discursiva opaca, donde el origen, procesamiento y circulación del mensaje escapan al escrutinio público y al control institucional.

Respecto al principio de equidad comunicativa, ninguno de los textos analizados incorpora una definición sustantiva del concepto, ni mecanismos explícitos para garantizar una distribución justa. La segmentación

personalizada basada en perfiles algorítmicos, hoy ampliamente utilizada en campañas comerciales y políticas, no encuentra límites normativos sólidos que prevengan la discriminación comunicativa. Si bien el RGPD establece restricciones frente a decisiones automatizadas basadas en categorías sensibles (como religión, orientación sexual o etnia), en la práctica estas restricciones resultan fácilmente eludibles mediante criterios opacos de optimización.

Tampoco existen normas que obliguen a realizar auditorías de equidad comunicativa, entendida como análisis de los efectos diferenciales de la comunicación automatizada sobre distintos grupos sociales, como puede observarse en la tabla 3. En este sentido, el marco normativo vigente reproduce una concepción neutralista de la tecnología, en la que la equidad se presupone como resultado espontáneo del cumplimiento procedimental, sin considerar los sesgos estructurales del diseño algorítmico ni sus efectos distributivos.

El análisis del discurso institucional presente en documentos como la ENIA (2020) muestra una retórica a favor de la “IA confiable, ética y centrada en el ser humano”, pero sin traducir estos principios en mecanismos de verificación o exigibilidad normativa. Esta retórica refuerza la figura del “consumidor ético informado” como garante último de la equidad, desplazando la responsabilidad estructural hacia el usuario y limitando el alcance de la ética pública digital a la buena voluntad de los actores tecnológicos.

Tabla 3

Evaluación normativa de los principios de transparencia, trazabilidad y equidad comunicativa

Norma analizada	Transparencia	Trazabilidad	Equidad comunicativa	Evaluación crítica global
AI Act	Sí (etiquetado de IA generativa), sin estándares definidos	Parcial (obligación de documentación técnica)	No aborda segmentación ni efectos diferenciados	Formaliza derechos sin condiciones sociocomunicativas de ejecución

Norma analizada	Transparencia	Trazabilidad	Equidad comunicativa	Evaluación crítica global
DSA	Sí (publicidad, segmentación, anunciante)	Parcial (plataformas deben explicar recomendaciones)	No regula la discriminación en publicidad automatizada	Aporta claridad procesual, pero deja vacíos semióticos
RGPD	Sí (información al usuario, art. 13–15)	Sí (registro de tratamiento de datos)	Parcial (protección contra decisiones automatizadas)	Centrado en datos, no en mensajes; protección individual
LOPDGDD	Sí (derechos digitales reforzados)	No especifica sobre contenidos automatizados	No aborda segmentación comunicativa	Refuerza RGPD, pero carece de instrumentos simbólicos
LSSI	No aborda automatización	No	No	Desactualizada frente a IA generativa
Ley General de Publicidad	Parcial (obligación de identificación de publicidad)	No	No	No prevé automatización ni personalización discursiva
ENIA (2020)	Retórica ética general	No operativo	Retórica, sin mecanismos de verificación	Declarativa, sin efecto normativo

Nota. Elaboración propia.

Podemos constatar, pues, que el fenómeno de la inteligencia artificial generativa se encuentra regulado de forma fragmentaria, asimétrica y conceptualmente insuficiente. En primer lugar, la caracterización de la automatización discursiva muestra que, si bien los textos normativos europeos más recientes, como el AI Act y la DSA, comienzan a reconocer la existencia de sistemas generativos y su relevancia comunicativa, la mayoría de las normas todavía se inscriben en un marco de interpretación centrado en datos personales, no en producción simbólica. La mayoría de las leyes españolas analizadas (como la LSSI o la Ley General de Publicidad) continúan operando bajo presupuestos prealgorítmicos, sin integrar conceptualmente la figura del emisor algorítmico ni los mecanismos de personalización automatizada del discurso.

En segundo lugar, en relación con la capacidad configuradora de gobernanza algorítmica, se observa una tensión entre intención reguladora y efectividad institucional. Las normas europeas avanzan en la creación de principios y obligaciones para las plataformas, pero dejan amplios márgenes

de discrecionalidad en la implementación, delegando responsabilidades clave en los propios actores privados. Por su parte, el marco normativo español, con excepción parcial de la LOPDGDD, carece de instrumentos estructurantes, lo que refuerza un modelo de gobernanza débil, fragmentado y reactivo, incapaz de anticipar los desafíos de la automatización simbólica.

Desde el punto de vista de los principios de transparencia, trazabilidad y equidad comunicativa se pone de relieve una importante brecha entre los valores declarados y los mecanismos normativos existentes para hacerlos efectivos. Aunque algunos textos, como el AI Act y la DSA, formulan obligaciones formales de transparencia y documentación, su vaguedad, su ambigüedad operativa y su desarticulación con sistemas de auditoría, rendición de cuentas o verificación social limitan su alcance. En particular, la ausencia casi total de instrumentos legales que garanticen la equidad comunicativa o la protección contra la discriminación simbólica confirma que el modelo regulador vigente prioriza el cumplimiento procedimental por encima de la justicia discursiva.

Discusión y conclusiones

Los resultados de este estudio evidencian que el uso de inteligencia artificial generativa (IAG) en la comunicación institucional y publicitaria está transformando los marcos simbólicos, normativos y culturales que regulan la producción de sentido en el espacio público. Más allá del fenómeno técnico, la automatización discursiva constituye una forma específica de reorganización del poder comunicativo, que afecta tanto a la legitimidad de los emisores institucionales como a la percepción de veracidad por parte de las audiencias (Fuchs, 2022; Striphas, 2015).

En consonancia con las perspectivas críticas sobre capitalismo digital, este trabajo ha confirmado que la IAG no puede ser entendida únicamente como una herramienta de eficiencia comunicativa, sino como un agente simbólico que interviene activamente en la configuración de subjetividades, identidades y emociones. Los hallazgos teóricos y empíricos apuntan a que la proliferación de discursos automatizados genera una despersonalización de la

comunicación institucional, que pone en crisis las categorías tradicionales de autoría, intencionalidad y responsabilidad (Gillespie, 2018; Zuboff, 2019).

Uno de los elementos clave detectados en los resultados es el vacío normativo que persiste en relación con la producción automatizada de mensajes persuasivos. Aunque normativas como el AI Act y la Digital Services Act (DSA) comienzan a incorporar obligaciones de transparencia y etiquetado, su implementación carece de una estructura robusta que garantice una trazabilidad efectiva y una equidad comunicativa verificable. En esta línea, autores como Mittelstadt et al. (2019) advierten que la transparencia tecnológica no puede reducirse a la disponibilidad formal de información, sino que debe estar orientada a la inteligibilidad social y a la explicabilidad contextual del mensaje automatizado.

Asimismo, el estudio ha mostrado que el enfoque predominante en los marcos normativos europeos sigue anclado en una lógica centrada en el individuo y en la protección de datos personales, como se refleja en el RGPD. Esta perspectiva legalista, basada en el consentimiento, la privacidad y la intervención individual, resulta insuficiente para abordar los impactos estructurales y simbólicos de la automatización del discurso en contextos institucionales.

Otra dimensión central en la discusión es la cuestión de la confianza pública en los mensajes automatizados. La investigación confirma que la legitimidad de la comunicación institucional no depende solo del contenido del mensaje, sino de su contexto de producción, de la percepción de intencionalidad y de la capacidad del receptor para interpretar críticamente su origen. Liao y Sundar (2022) proponen que la confianza en sistemas de IA debe entenderse como una construcción relacional, basada en motivación, control, transparencia y apoyo interpretativo. En este sentido, etiquetar un mensaje como “generado por IA” resulta ineficaz si no va acompañado de dispositivos institucionales que garanticen la inteligibilidad funcional y ética del proceso.

Asimismo, Jacovi et al. (2020) sostienen que la confianza tecnológica opera como una forma de contrato implícito entre el usuario y el sistema, en el que la expectativa de predictibilidad y explicabilidad juega un papel central. Desde esta óptica, el discurso automatizado en entornos corporativos debe evaluarse no solo por su adecuación legal, sino por su capacidad para integrarse en marcos socioculturales de interpretación legítima. Cuando esta integración falla, por ejemplo, en contextos donde no se informa adecuadamente sobre la naturaleza generativa del mensaje, se erosiona la percepción de autenticidad institucional y se debilita la confianza pública.

En relación con la cultura algorítmica, los resultados respaldan la idea de que los sistemas generativos no solo producen contenido, sino que instauran nuevas lógicas de visibilidad, atención y relevancia cultural (Striphas, 2015). Este cambio afecta particularmente al ámbito de la comunicación institucional, donde la delegación discursiva en sistemas de IA puede socavar las estructuras simbólicas tradicionales sobre las que se sostiene la credibilidad pública. La automatización del mensaje no es neutral: implica una redistribución de la agencia comunicativa, un desplazamiento del locus de la autoría y una reconfiguración del vínculo entre emisor y audiencia (Danesi, 2022; Nass & Moon, 2000).

La dimensión regulatoria española, representada por la Ley General de Publicidad, la LSSI o la LOPDGDD, muestra una desconexión notable con la evolución tecnológica de la comunicación digital. Mientras las normas europeas comienzan a articular principios orientadores, las leyes españolas mantienen un enfoque anclado en presupuestos prealgorítmicos, sin contemplar la figura del emisor no humano ni los efectos diferenciados de la automatización simbólica (Domínguez, 2023; Peña, 2021). Este desfase entre avance tecnológico y respuesta jurídica contribuye a consolidar un régimen de comunicación automatizada sin contrapesos públicos efectivos, lo que incrementa las asimetrías de poder entre plataformas tecnológicas, organizaciones emisoras y audiencias.

Los resultados también sugieren que la transparencia formal no equivale a transparencia significativa. El etiquetado automático de mensajes

generados por IA, si no se acompaña de mecanismos comprensibles y contextuales de explicación, puede generar efectos adversos, como la pérdida de confianza o el rechazo emocional. Como indican Bajaña-Cedeño et al. (2025), la sobreexposición de la etiqueta “contenido generado por IA” puede ser contraproducente si no va acompañada de pedagogías comunicativas que doten de sentido a la distinción entre humano y algorítmico en contextos de interacción institucional.

Frente a este escenario, el presente artículo propone una doble línea de acción. En primer lugar, es imprescindible reformular los fundamentos normativos que rigen la comunicación automatizada, incorporando principios de equidad simbólica, trazabilidad discursiva e inteligibilidad pública como ejes vertebradores de la regulación de sistemas generativos. En segundo lugar, resulta urgente reforzar la capacidad institucional para auditar, verificar y sancionar los usos indebidos de la automatización simbólica, especialmente en campañas dirigidas a públicos vulnerables o en contextos de desinformación. Ello implica el desarrollo de infraestructuras de gobernanza que no solo regulen la técnica, sino que interpreten críticamente su inserción en el campo de lo social, lo comunicativo y lo político.

Referencias

- Aminahuel, A., & Rodríguez, M. (2024). Gobernanza de plataformas digitales y de inteligencia artificial: Contradicciones y concentraciones del poder tecnológico. *Temas y Problemas de Comunicación*, 22, 31-46. <https://www2.hum.unrc.edu.ar/ojs/index.php/TyPC/article/view/2095>
- Arsenal, L., & López, R. G. (2025). *Democracia 4.0: IA y automatismos para la futura gobernanza*. Kokapeli Ediciones.
- Bajaña-Cedeño, D., Freixa, P., & Codina, L. (2025). *Transparencia y responsabilidad en el uso de la IA en medios de comunicación: Análisis comparativo* (Informe No. 2). Universitat Pompeu Fabra, Departament de Comunicació. <https://doi.org/10.31009/coc.02.2025>
- Benito-García, J. M. (2025). Mercado audiovisual e inteligencia artificial. Relaciones y aspectos éticos. *European Public & Social Innovation Review*, 10, 1-17. <https://doi.org/10.31637/epsir-2025-2156>

- Bijker, W. E., Hughes, T. P., & Pinch, T. (Eds.). (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. The MIT Press.
- Borrás, S., & Edler, J. (Eds.). (2020). *The governance of socio-technical systems: Explaining change*. Edward Elgar Publishing. <https://doi.org/10.4337/9781784710194>
- Cabrero, J. D. B., & Caldevilla-Domínguez, D. (2025). Relaciones públicas frente a la IA: retos, oportunidades y consideraciones. *European Public & Social Innovation Review*, 10, 1-22. <https://doi.org/10.31637/epsir-2025-2144>
- Danesi, C. (2022). *El imperio de los algoritmos: LA inclusiva, ética y al servicio de la humanidad*. Editorial Galerna.
- Domínguez, A. G. (2023). Las exigencias de transparencia para los sistemas algorítmicos de recomendación, selección de contenidos y publicidad en línea en el nuevo Reglamento Europeo de Servicios Digitales. *Revista española de la transparencia*, 17(1), 113-130. <https://doi.org/10.51915/ret.309>
- España. (1988, 11 de noviembre). Ley 34/1988, de 11 de noviembre, General de Publicidad. *Boletín Oficial del Estado*, 274. <https://www.boe.es/eli/es/l/1988/11/11/34>
- España. (2002, 12 de julio). Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico. *Boletín Oficial del Estado*, 166, 25388–25403. <https://www.boe.es/eli/es/l/2002/07/11/34>
- España. (2018, 6 de diciembre). Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. *Boletín Oficial del Estado*, 294. <https://www.boe.es/eli/es/lo/2018/12/05/3>
- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *methaodos. Revista de Ciencias Sociales*, 11(2), m231102a10. <https://doi.org/10.17502/mrcs.v11i2.710>
- Fuchs, C. (2022). *Digital capitalism: Media, communication and society* (Vol. 3). Routledge. <https://doi.org/10.4324/9781003222149>
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press. <https://doi.org/10.12987/9780300235029>
- Hildebrandt, M. (2022). *Law for computer scientists and other folk*. Oxford University Press. <https://doi.org/10.1093/oso/9780198860877.001.0001>
- Hinojosa-Becerra, M., Marín-Gutiérrez, I., & Maldonado-Espinosa, M. (2024). Inteligencia artificial y la producción audiovisual. En *Espejo de monografías de comunicación social* (Vol. 23, pp. 117–139). <https://doi.org/10.52495/c6.emcs.23.ti12>
- Hueso, L. C., & Claramunt, C. (2022). Transparencia y explicabilidad de la inteligencia artificial y “compañía” (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad...): Para qué, para

- quién y cuánta. En *Transparencia y explicabilidad de la inteligencia artificial* (pp. 29-70).
- Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 624–634). <https://doi.org/10.1145/3442188.3445923>
- Jasanoff, S. (Ed.). (2004). *States of knowledge: The co-production of science and the social order*. Routledge.
- Liao, Q. V., & Sundar, S. S. (2022). Designing for responsible trust in AI systems: A communication perspective. En *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1257-1268). <https://doi.org/10.1145/3531146.3533182>
- Macías, J. I. Z. (2025). Narrativas del futuro: Inteligencia artificial en la comunicación visual. *Revista Enfoques de la Comunicación*, 13, 190-225.
- Martínez, K. C., Rodríguez, J. A. A., & Rentería, A. S. M. (2024). Desafíos éticos de la inteligencia artificial generativa en las nuevas formas organizacionales. *Revista Digital de Tecnologías Informáticas y Sistemas*, 8(1), 46-55. <https://doi.org/10.61530/redtis.vol8.n1.2024.209>
- Mittelstadt, B. D., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. En *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 279–288). <https://doi.org/10.1145/3287560.3287574>
- Montero, M. (2018). *Análisis sociológico de la regulación digital*. Editorial CIS.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Peña-Fernández, S., Peña-Alonso, U., & Eizmendi-Iraola, M. (2023). El discurso de los periodistas sobre el impacto de la inteligencia artificial generativa en la desinformación. *Estudios sobre el Mensaje Periodístico*, 29(4), 833-841. <https://doi.org/10.5209/esmp.88673>
- Peña, J. C. H. (2021). Gobernanza de la inteligencia artificial en la Unión Europea. La construcción de un marco ético-jurídico aún inacabado. *Revista General de Derecho Administrativo*, 56.
- Pérez-Ugena Coromina, M. (2024). Protección de datos en la aplicación de sistemas de inteligencia artificial. *Revista Anuario Parlamento y Constitución*, 25, 207–232. <https://doi.org/10.71206/rapc.393>
- Piedra Alegría, J. (2023). Anotaciones iniciales para una reflexión ética sobre la regulación de la Inteligencia Artificial en la Unión Europea. *Revista de Derecho (Universidad Católica Dámaso A. Larrañaga, Facultad de Derecho)*, 28. <https://doi.org/10.22235/rd28.3264>

- Piedra Alegría, J. (2024). Democracias generativas: Inteligencia artificial y manipulación en el siglo XXI. *Trayectorias Humanas Trascontinentales*, 18. <https://doi.org/10.25965/trahs.6334>
- Pierson, P. (1993). When effect becomes cause: Policy feedback and political change. *World Politics*, 45(4), 595–628. <https://doi.org/10.2307/2950710>
- Poveda, J. F. S., & Saltos, G. D. C. (2024). Algoritmos en redes sociales y estrategias comunicacionales para el posicionamiento de medios tradicionales. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(6), 1670–1686. <https://doi.org/10.56712/latam.v5i6.3114>
- Quan-Haase, A. (2013). *Technology and society: Social networks, power, and inequality*. Oxford University Press.
- Rodríguez, D. F. (2024). Gobernanza y Reglamento de Inteligencia Artificial desde la primera óptica de OpenAI. *Derecho & Sociedad*, 63, 243–258. <https://doi.org/10.18800/dys.202402.016>
- Santana Ramos, E. M. (2024). El futuro de la inteligencia artificial en el marco europeo. *Revista Brasileira de Políticas Públicas*, 14(2). <https://doi.org/10.5102/rbpp.v14i2.9572>
- Seijas, R. (2023). Políticas de comunicación e inteligencia artificial: nuevos desafíos. URVIO. *Revista Latinoamericana de Estudios de Seguridad*, 37, 45–62. <https://doi.org/10.17141/urvio.37.2023.5992>
- Striphas, T. (2015). Algorithmic culture. *European Journal of Cultural Studies*, 18(4-5), 395–412. <https://doi.org/10.1177/1367549415577392>
- Unión Europea. (2016, 27 de abril). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento general de protección de datos). *Diario Oficial de la Unión Europea*, L 119, 1–88. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- Unión Europea. (2022, 19 de octubre). Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo, de 19 de octubre de 2022, relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales). *Diario Oficial de la Unión Europea*, L 277, 1–102. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>
- Unión Europea. (2024, 13 de junio). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2019/2144 y (UE) 2020/741 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). *Diario Oficial de la Unión Europea*, L 206, 1–101. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

- Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97-112. <https://doi.org/10.9785/cr-2021-220402>
- Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4), 505–523. <https://doi.org/10.1111/rego.12158>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Liderazgo escolar, Inteligencia Artificial y el imperativo ético en el enfoque STEM

School leadership, Artificial Intelligence, and the ethical imperative in the STEM approach

Johanna Solange Godoy Cerda

johanna.godoy2021@umce.cl

Universidad Metropolitana de Ciencias de la Educación

ORCID ID 0009-0003-9500-4986

Resumen: La integración de la Inteligencia Artificial (IA) en la educación, en el marco del enfoque STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), plantea un desafío crucial que trasciende la mera optimización técnica para avanzar hacia la justicia educativa. Esta investigación cualitativa tuvo como objetivo comprender cómo el liderazgo escolar y la ética en el uso de la IA inciden en la implementación significativa y transformadora del enfoque STEM en instituciones de educación secundaria. Se aplicaron entrevistas semiestructuradas a directivos de establecimientos con trayectorias en innovación. Los resultados revelan un conocimiento incipiente sobre la dimensión pedagógica y ética de la IA, lo que limita su potencial transformador. Sin embargo, se identifican factores críticos que deben ser considerados para el éxito de la implementación de STEM: el liderazgo distribuido, la cultura colaborativa docente y un compromiso ético explícito. Se concluye que el liderazgo escolar es determinante para incorporar y orientar la IA al servicio de objetivos transformadores como el desarrollo del pensamiento crítico y la reducción de brechas estructurales con el fin de garantizar que la enseñanza STEM se convierta en una herramienta de justicia social y equidad.

Palabras clave: Educación STEM; ética tecnológica; justicia social; inteligencia artificial; liderazgo educativo.

Abstract: The integration of Artificial Intelligence (AI) into education, within the framework of the STEM (Science, Technology, Engineering, and Mathematics) approach, poses a crucial challenge that transcends mere technical optimization to advance educational justice. This qualitative research aimed to understand how school leadership and ethics in the use of AI impact the meaningful and transformative implementation of the STEM approach in secondary education institutions. Semi-structured interviews were conducted with school principals with a history of innovation. The results reveal an incipient understanding of the pedagogical and ethical dimensions of AI, which limits its transformative potential. However, critical factors are identified that must be considered for the successful implementation of STEM: distributed leadership, a collaborative teaching culture, and an explicit ethical commitment. It is concluded that school leadership is crucial in incorporating and guiding AI in the service of transformative objectives such as developing critical thinking and reducing structural gaps, ensuring that STEM education becomes a tool for social justice and equity.

Keywords: STEM education; technological ethics; social justice; artificial intelligence; educational leadership.

Introducción

La educación contemporánea se encuentra en una encrucijada a nivel global, obligada a equilibrar la necesidad de preparar a los estudiantes para las competencias críticas de una sociedad impulsada por la tecnología y el imperativo ético de mitigar las desigualdades sociales que persisten en los sistemas escolares. En este contexto, el enfoque STEM (Science, Technology, Engineering, and Mathematics) ha trascendido a su propósito inicial de desarrollar una fuerza laboral técnica (Zhan & Niu, 2023) para consolidarse como una plataforma que permite avanzar hacia la justicia educativa. El valor real de STEM reside en su potencial para fomentar la alfabetización crítica, permitiendo a los estudiantes interactuar con la ciencia y la tecnología desde

una postura ética y socialmente responsable (Barton & Tan, 2010; Calabrese, 2002).

La irrupción de la Inteligencia Artificial (IA), particularmente la IA generativa, constituye un punto de inflexión pedagógica, puesto que promete redefinir el diseño instruccional mediante herramientas de personalización del aprendizaje y optimización de la retroalimentación a una escala sin precedentes (Barkoczi et al., 2024). Sin embargo, este potencial transformador está intrínsecamente ligado a una tensión ética crucial que señala que, si bien la IA es un motor potencial de equidad, su implementación apresurada o deficiente puede exacerbar las desigualdades estructurales. La opacidad algorítmica, la inherente reproducción de sesgos históricos en los datos de entrenamiento, y los desafíos en la gobernanza de la privacidad de datos (Miao et al., 2021) plantean el riesgo de que la IA se convierta en un factor de profundización de las brechas, especialmente en contextos socioeconómicos vulnerables.

Por consiguiente, la integración de la IA en el enfoque STEM no puede ser considerada un simple proceso técnico, sino un desafío socio-técnico y cultural que exige una gobernanza ética robusta a nivel local. Es en este punto de convergencia donde el liderazgo escolar emerge como la variable crítica y el arquitecto cultural de la transformación. El director debe trascender desde la figura de administrador para asumir un rol de líder transformador y ético (Shields, 2017), capaz de actuar como mediador principal entre las directrices tecnológicas globales y las necesidades pedagógicas situadas de su comunidad (Harris & Jones, 2023).

La principal brecha identificada para esta implementación significativa no es tecnológica, sino que se establece en la comprensión profunda y la gestión ética por parte de los equipos directivos. El éxito, por lo tanto, depende de la capacidad de los líderes para movilizar a sus comunidades hacia un uso de la tecnología que esté alineado con los principios de equidad y el bien común.

Dado lo anterior es relevante comprender ¿De qué manera el liderazgo de los directores escolares, junto al uso ético de la Inteligencia

Artificial, incide en la implementación significativa del enfoque STEM en instituciones de educación secundaria, considerando las oportunidades, obstaculizadores y condiciones institucionales asociadas?

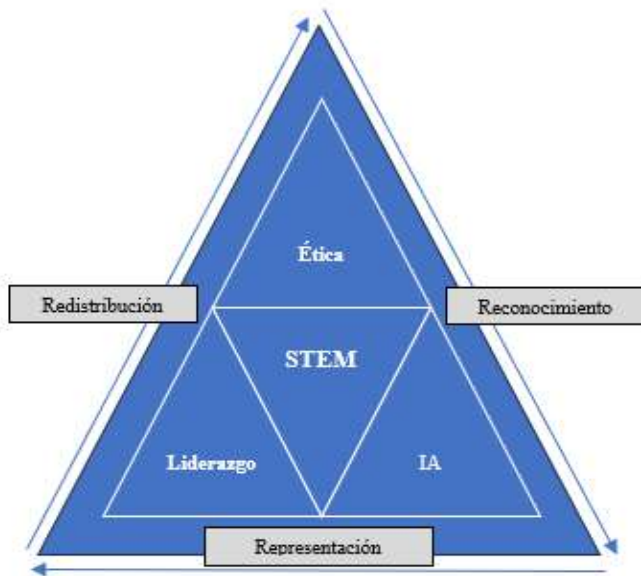
Fundamentación

El marco conceptual de esta investigación se fundamenta en la concepción de STEM como una herramienta de Justicia Social, rechazando la neutralidad de la ciencia y la tecnología. Siguiendo a Barton y Tan (2010), la educación científica debe ser entendida como un espacio de emancipación, reflexión y acción social, donde los estudiantes se involucren activamente en la solución de problemas auténticos de sus comunidades, conectando el currículo con la realidad sociopolítica. Esta visión se alinea con la perspectiva que concibe la educación STEM como un derecho civil que debe asegurar la excelencia y la equidad para todos los grupos demográficos (Tate, 2001).

Se adopta la perspectiva de Justicia Educativa con enfoque tridimensional de Fraser (2008), el cual es crucial para evaluar la implementación de STEM y el uso de IA. Podemos observar esta perspectiva en la Figura 1.

Figura 1

Justicia Educativa según Nancy Fraser



Nota. Elaboración propia.

La Figura 1 presenta el pensamiento de Fraser (2008) al respecto de la Justicia Educativa, la cual se establece en base a tres dimensiones:

- a) **Redistribución:** Se refiere a un acceso equitativo a los recursos tecnológicos y pedagógicos (laboratorios, hardware, software de IA). A partir del punto anterior se puede establecer que la ausencia de la redistribución perpetúa significativamente la brecha digital entre establecimientos.
- b) **Reconocimiento:** Implica la validación de conocimientos y experiencias de grupos históricamente marginados. La implementación de un enfoque STEM basado en la justicia y el reconocimiento promueve la integración de saberes locales o

indígenas en el diseño de soluciones tecnológicas (Calabrese, 2002).

- c) Representación: Asegura la participación efectiva de los estudiantes y las comunidades en la definición de los problemas y soluciones, garantizando que el diseño curricular de STEM esté influenciado por la diversidad de la población escolar.

La Inteligencia Artificial, integrada a STEM, debe ser evaluada sistémicamente no solo por su eficiencia pedagógica, sino por cómo amplía o restringe estas tres dimensiones de la justicia educativa, asegurando que la tecnología no sea solo consumida, sino analizada y co-creada por estudiantes con pensamiento crítico.

Inteligencia artificial: La tensión socio-técnica y el imperativo ético

La IA se enmarca en esta investigación como un sistema socio-técnico complejo (Selwyn, 2024), donde el software y los algoritmos no operan en un vacío, sino que son el resultado de las estructuras sociales, culturales y políticas incrustadas en sus datos. Esta perspectiva subraya que las fallas de la IA no son meramente técnicas, sino que son el reflejo de injusticias sociales preexistentes.

El riesgo central radica en la opacidad algorítmica y el sesgo inherente, el cual se manifiesta cuando los datos utilizados para entrenar modelos de aprendizaje automatizado reproducen patrones de discriminación de género, raza o nivel socioeconómico (Boateng & Boateng, 2025). Este sesgo, se presenta al automatizar decisiones educativas (ej. tracking de estudiantes o recomendaciones de rutas de aprendizaje), lo que representa una amenaza directa y sistémica a la equidad.

El Imperativo Ético se sustenta en la Recomendación de la UNESCO sobre la Ética de la IA (Miao et al., 2021), que exige un enfoque centrado en los derechos humanos y el desarrollo sostenible especialmente los Objetivos de Desarrollo Sostenible 4, 5 y 10. La adopción significativa de la IA requiere

superar la alfabetización digital básica e impulsar la Alfabetización Crítica de la IA. Esto implica tres puntos críticos, los cuales son descritos en la Tabla 1.

Tabla 1

Puntos Críticos del Imperativo Ético

Puntos Críticos del Imperativo Ético	
Puntos Críticos	Descripción
1) Transparencia y Explicabilidad	Entender cómo y por qué la IA toma ciertas decisiones.
2) Fomento del Pensamiento Crítico	Usar la IA no para la automatización, sino para la reflexión sobre su impacto social y sus limitaciones (Southgate, 2020).
3) Mitigación de Riesgos	Desarrollar marcos éticos claros a nivel institucional que orienten el uso docente, aseguren la privacidad de los datos de los estudiantes y mantengan un enfoque de "Human in the Loop", donde el docente y el estudiante preserven la autonomía y la capacidad de cuestionamiento (Chen, 2024).

Nota. Elaboración propia.

La Tabla 1 sobre Puntos Críticos del Imperativo Ético sintetiza tres dimensiones fundamentales para orientar el uso responsable de la inteligencia artificial (IA) en contextos educativos. En primer lugar, se enfatiza la transparencia y explicabilidad, entendida como la necesidad de comprender cómo y por qué los sistemas de IA generan determinados resultados, lo que favorece la confianza y la apropiación crítica de estas herramientas. En segundo lugar, se destaca el fomento del pensamiento crítico, promoviendo que la IA sea utilizada no como un mecanismo de automatización acrítica, sino como un recurso para cuestionar sus propios alcances, limitaciones e implicaciones sociales (Southgate, 2020). Finalmente, se aborda la mitigación de riesgos, lo que implica la creación de marcos éticos claros a nivel institucional que regulen su implementación, protejan la privacidad de los datos estudiantiles y mantengan un enfoque de Human in the Loop. Este último resguarda la centralidad del docente y del estudiante en el proceso de toma de decisiones, asegurando tanto la autonomía como la capacidad reflexiva frente a la tecnología (Chen, 2024).

Liderazgo escolar transformador y distribuido

El liderazgo escolar es el catalizador necesario para traducir los principios de justicia y ética en acciones pedagógicas concretas. Se fundamenta en el concepto de Liderazgo Transformador (Leithwood & Sun, 2012; Shields, 2017), que va más allá de la gestión administrativa o transaccional para centrarse en la visión compartida y la justicia social como ejes del cambio cultural.

En el contexto tecnológico, este liderazgo se manifiesta como Liderazgo Tecno-Ético y Distribuido (Spillane, 2005). El liderazgo efectivo en la era de la IA debe poseer las siguientes cualidades, vitales para el éxito del enfoque STEM:

- **Visión Crítica y Prospectiva:** Capacidad para discernir entre la innovación genuina y el mero hype tecnológico, orientando la IA hacia objetivos pedagógicos profundos y no a la simple eficiencia. El líder debe ser un aprendiz constante de la tecnología y la ética.
- **Liderazgo Distribuido y Agencia Docente:** El fomento de la autonomía docente y la cultura colaborativa (Harris & Jones, 2023; Spillane, 2005) es esencial. La IA es demasiado compleja para ser gestionada por un solo individuo; su uso ético requiere que los profesores desarrollen agencia para experimentar, evaluar críticamente y adaptar las herramientas de IA a sus aulas de manera segura.
- **Liderazgo Situado y Sensibilidad Contextual:** Implica la capacidad de adecuar las políticas tecnológicas globales y los marcos éticos a las necesidades, desafíos y contextos específicos de la comunidad escolar. Este liderazgo asegura que la IA sirva al bien común de esa población estudiantil particular, mitigando activamente las desigualdades que son parte de ese entorno.

Metodología

Diseño y enfoque

Esta investigación se erige sobre el paradigma interpretativo constructivista, el cual dicta una postura que reconoce la realidad como una entidad plurivocal y subjetivamente construida en la interacción social. Desde esta óptica, la epistemología se orienta hacia el proceso de Verstehen (comprensión profunda), donde el conocimiento es co-creado entre el investigador y los participantes. El foco no es la causalidad o la generalización estadística, sino la captura significativa de los marcos interpretativos que los líderes escolares utilizan para dar sentido a la irrupción de la Inteligencia Artificial (IA) en el ámbito STEM.

Se adoptó el enfoque cualitativo, considerado indispensable para dilucidar en detalle las percepciones, la configuración de estrategias de liderazgo y la complejidad de los dilemas éticos que son inherentemente subjetivos, contextuales y, por ende, inasibles para las aproximaciones cuantitativas.

El diseño metodológico específico es de naturaleza exploratorio-descriptiva. Este diseño fue seleccionado por su rigor pragmático para describir y conceptualizar constructos y procesos como la gestión institucional en la intersección de la IA y STEM en un contexto educativo singular (la educación secundaria chilena). El fenómeno abordado es emergente y ha tenido una rápida mutación, por este motivo la evidencia empírica situada y local es escasa, por este motivo generar un conocimiento inductivo de alto valor heurístico y analítico es vital para futuras intervenciones.

Participantes y muestreo

Se utilizó un muestreo intencionado bajo la modalidad de criterio, enfocado en la selección de directivos que tuvieran un conocimiento y experiencia demostrables en la gestión de la innovación educativa y el enfoque STEM.

El tamaño final de la muestra fue de diez (N=10) directivos de educación secundaria en Chile. El proceso de muestreo se detuvo una vez alcanzado el criterio de saturación teórica, momento en que la nueva información recopilada dejó de generar categorías, temas emergentes o variaciones significativas en los constructos ya definidos, asegurando la suficiencia y riqueza de los datos.

Criterios de inclusión

Los participantes fueron seleccionados basándose en los criterios de inclusión que presenta a continuación la Tabla 2.

Tabla 2

Criterios de Inclusión

Criterios de Inclusión	
Criterio de inclusión	Descripción
1) Cargo Directivo	Ser director activo en un establecimiento de educación secundaria en Chile.
2) Experiencia	Contar con al menos tres años de experiencia continua en el cargo directivo.
3) Innovación Educativa	El establecimiento debe poseer un Proyecto Educativo Institucional (PEI) que incorpore explícitamente el enfoque STEM y/o la integración tecnológica avanzada, demostrando un compromiso institucional con la innovación.

Nota. Elaboración propia.

La Tabla 2 presenta los criterios de inclusión con los cuales fueron seleccionados los participantes de esta investigación. A partir de estos criterios la muestra final incluyó una diversidad en la dependencia administrativa (pública y particular subvencionada) y una distribución territorial representativa de diferentes realidades socioeconómicas dentro de la Región Metropolitana, con el fin de obtener una lectura contextualizada de las oportunidades y obstaculizadores para abordar la dimensión de justicia educativa.

El contacto con los participantes se realizó a través de tres fases:

- 1) Identificación inicial, momento en el cual se contactó a redes de colaboración académica de directivos escolares para generar una lista preliminar de candidatos que cumplieran con el criterio de innovación educativa.
- 2) Verificación de criterios, realizado a través del envío de un formulario de autoevaluación inicial para confirmar el criterio de cargo y experiencia.
- 3) Invitación formal, en esta instancia los candidatos preseleccionados recibieron una invitación formal por correo electrónico con una descripción detallada de la investigación, el propósito de la entrevista, junto con el compromiso de confidencialidad y anonimización. Luego de la aceptación de esta invitación se procedió a programar las entrevistas.

Instrumento y recolección de datos

La información se recopiló a través de entrevistas semiestructuradas individuales, una técnica que permite la flexibilidad necesaria para explorar en profundidad los significados y las narrativas de los directivos, manteniendo al mismo tiempo un protocolo temático uniforme que garantiza la comparabilidad de los datos.

El Protocolo de Entrevista se estructuró en tres grandes dominios temáticos, directamente alineados con el objetivo central de la investigación, el detalle se presenta a continuación en la Tabla 3.

Tabla 3

Dominios Temáticos

Dominios Temáticos	
Dominios Temáticos	Descripción
Dominio I Conocimiento y Apropiación de IA	Indagación sobre la familiaridad con la IA generativa, su uso personal/profesional y la percepción de su impacto pedagógico.

Dominios Temáticos	
Dominios Temáticos	Descripción
Dominio II Estrategias de Liderazgo y Condiciones Institucionales	Exploración de las estrategias utilizadas para fomentar la cultura STEM, la autonomía docente y la gestión de recursos tecnológicos.
Dominio III Ética y Justicia Educativa	Reflexión sobre los dilemas de equidad, el sesgo algorítmico y los marcos éticos desarrollados para la integración de la IA.

Nota. Elaboración propia.

La Tabla 3 nos presenta en detalle los dominios temáticos utilizados en las entrevistas, cada entrevista tuvo una duración promedio de 60 a 75 minutos. Las entrevistas fueron realizadas en modalidad virtual, luego fueron grabadas digitalmente (previo consentimiento informado de los participantes) y posteriormente transcritas. Se garantizó la confidencialidad mediante la asignación de códigos alfanuméricos a cada participante.

Procedimiento de análisis de datos

El análisis de las transcripciones se realizó siguiendo el método de Análisis Temático sistemático propuesto por Braun y Clarke (2006) una metodología flexible y rigurosa que permite identificar, analizar y reportar temas presentes en los datos recolectados.

Se utilizó el software de análisis cualitativo ATLAS.ti para gestionar, organizar y codificar los datos de manera eficiente. El análisis de datos se ejecutó en seis fases iterativas, las cuales se presentan a continuación en la Tabla 4.

Tabla 4

Fases de Análisis de Datos

Fases de Análisis de Datos	
Fases de Análisis	Descripción
1. Familiarización con los datos	Lectura repetida y activa de todas las transcripciones y toma de notas iniciales sobre ideas recurrentes y de interés.
2. Generación de Códigos Iniciales	Identificación de fragmentos de texto relevantes (citas) y etiquetado con códigos descriptivos de primer orden (ej. "temor al sesgo", "liderazgo

Fases de Análisis de Datos	
Fases de Análisis	Descripción
(Codificación Abierta)	descentralizado", "falta de formación ética"). Estos códigos se mantuvieron cercanos al lenguaje de los participantes.
3. Búsqueda de Temas	Agrupación y clasificación de los códigos iniciales en temas potenciales más amplios y abstractos (ej. Obstaculizadores de la IA, Roles del Liderazgo Ético).
4. Revisión de Temas	Refinamiento, fusión o eliminación de temas. Se verificó la coherencia interna de los datos dentro de cada tema y la distinción externa entre los temas para asegurar que capturen las esencias narrativas. El proceso incluyó la fusión, escisión o eliminación de temas según la saturación de los datos. Se utilizaron mapas temáticos para visualizar la estructura conceptual y las relaciones jerárquicas entre códigos y temas.
5. Definición y Nombramiento de Temas	Creación de definiciones claras y concisas para cada tema final (que se convierten en las categorías de resultados) y selección de un nombre que refleje fielmente su contenido.
6. Producción del Informe	La fase final consistió en integrar el análisis temático con el marco teórico y los objetivos del estudio. Se utilizó evidencia textual (citas) para sustentar la validez de los temas y se estableció una conexión explícita entre las categorías emergentes y la pregunta de investigación para construir la narrativa final de los resultados.

Nota. Elaboración propia.

La Tabla 4 nos presenta las fases de análisis de datos, dando cuenta de un proceso metódico que garantiza la credibilidad, la auditabilidad y el rigor de los hallazgos cualitativos presentados, estableciendo un vínculo sólido entre la voz de los participantes y las conclusiones académicas del estudio.

Rigor científico y consideraciones éticas

Para asegurar el rigor metodológico, se aplicaron los siguientes criterios:

- Credibilidad o validez interna, para garantizar esta validez se utilizó la triangulación de datos, mediante la comparación de las narrativas entre diferentes directivos y la verificación, supervisión y debate de los códigos iniciales para minimizar el sesgo interpretativo.

- Transferibilidad o validez externa, mediante la descripción detallada del contexto, en este caso educación secundaria en Chile, los criterios de inclusión y perfil de los participantes. esta información facilitara evaluar la pertinencia de transferir los hallazgos a otros contextos similares.
- Dependencia o fiabilidad, este criterio fue desarrollado mediante el uso de un protocolo de entrevista estandarizado y la documentación detallada del proceso de codificación en ATLAS.ti, lo que permite asegurar que el proceso analítico pueda ser auditado.

La investigación cumplió con principios éticos mediante el uso de:

- Consentimiento informado: Todos los participantes firmaron un formulario de consentimiento informado que detallaba el propósito de la investigación, el procedimiento, la duración de la entrevista, el uso de las grabaciones y su derecho a retirarse de la investigación en cualquier momento sin ninguna consecuencia asociada.
- Anonimización y confidencialidad: Se garantizó la absoluta confidencialidad de la información. Las transcripciones se codificaron con identificadores alfanuméricos (D01-D10) y todos los datos personales o institucionales sensibles que pudieran identificar a los participantes o sus establecimientos fueron omitidos en la transcripción final.
- Voluntariedad: La participación fue completamente voluntaria y no hubo compensación económica, asegurando que no existiera coacción en la toma de decisiones.

Resultados

Los hallazgos de la investigación se agruparon en tres categorías principales que reflejan los desafíos, las condiciones habilitantes y las exigencias

normativas en la intersección del liderazgo escolar, la ética de la IA y el enfoque STEM. Para clarificar la relación entre la evidencia (códigos) y los hallazgos (temas/categorías), se presenta un resumen de la estructura analítica, en la Tabla 5.

Tabla 5
Estructura Analítica de Categorías, Temas y Códigos Emergentes

Estructura Analítica de Categorías, Temas y Códigos Emergentes		
Fases de Análisis	Descripción	Códigos
1-. Conocimiento y Urgencia Formativa	Conocimiento Incipiente y Uso Básico El uso de IA se limita a herramientas generativas conocidas (ChatGPT) y funciones administrativas.	- Reconocimiento de marca (ChatGPT). - Usos administrativos básicos. - Ausencia de visión estratégica.
	Riesgos Ignorados Falta de protocolos y reflexión sistemática sobre los dilemas éticos y el sesgo.	- Ignorancia sobre sesgo algorítmico. - Ausencia de protocolos éticos. - Percepción de urgencia formativa (demanda de capacitación crítica).
2-. Condiciones Habilitantes para la Transformación.	Liderazgo Distribuido y Agencia Docente Los directivos actúan como facilitadores, otorgando autonomía al profesorado para innovar.	- Liderazgo descentralizado. - Espacios protegidos (Comités/Redes STEM). - Modelo bottom-up de innovación, en el cual las decisiones se construyen a partir de la experiencia, ideas y propuestas del equipo
	Cultura Colaborativa y Co-creación El aprendizaje de IA ocurre en espacios colectivos de debate ético y co-diseño pedagógico.	- Compartir prompts y secuencias didácticas. - Debates informales sobre dilemas. - Fortalecimiento de la Representación (Justicia Educativa).
3-. El Imperativo Ético del Uso de la IA.	Desafío de la Integridad Académica Preocupación por el uso no ético de la IA por parte de los estudiantes.	- Uso no ético en proyectos. - Necesidad de guías institucionales. - Tensión en la autenticidad de la evaluación.
	Gobernanza de Datos y Prioridad Pedagógica Demanda de transparencia en la gestión de datos y orientación a la enseñanza.	- Resguardo de privacidad de datos. - Resistencia a la tecnología. - Orientación a enriquecer el rigor cognitivo.

Nota. Elaboración propia.

La Tabla 5 presenta el análisis temático que logró la identificación de tres Categorías Principales. Estas categorías articulan el impacto del liderazgo escolar y la ética en la implementación de la Inteligencia Artificial (IA) en el enfoque STEM, y son: 1) Conocimiento y Urgencia Formativa, 2) Condiciones Habilitantes para la Transformación, y 3) El Imperativo Ético del Uso de la IA. La matriz analítica demuestra un proceso riguroso de condensación de datos, donde los códigos (ej. "Reconocimiento de marca", "Liderazgo descentralizado") se agruparon temáticamente (ej. "Conocimiento Incipiente y Uso Básico") para finalmente conformar categorías conceptuales más amplias.

En conjunto, estos hallazgos sugieren que la integración significativa de la IA trasciende la mera adquisición de herramientas. Mientras que el “Conocimiento y Urgencia Formativa” revela una brecha crítica en la comprensión estratégica de la IA y sus riesgos éticos, la categoría de “Condiciones Habilitantes” enfatiza que la innovación efectiva se logra mediante un liderazgo distribuido y una cultura colaborativa que fomenta la agencia docente. Por último, la categoría denominada “El Imperativo Ético” subraya la necesidad urgente de marcos de gobernanza y protocolos claros para abordar los desafíos de la integridad académica y la privacidad de datos, asegurando que la tecnología permanezca orientada a enriquecer el rigor cognitivo y la justicia educativa.

Discusión

La brecha de la alfabetización crítica y el riesgo de instrumentalización

Los resultados demuestran que el conocimiento incipiente y superficial de la IA por parte de los directivos es el principal nudo crítico que impide la transición hacia un liderazgo auténticamente transformador y tecno-ético.

Al no comprender las implicaciones del sesgo algorítmico (Boateng & Boateng, 2025) o las dinámicas de opacidad de los modelos generativos, el liderazgo reduce la IA a una herramienta de eficiencia administrativa o técnica, cayendo en la instrumentalización. Esta postura contrasta

directamente con la visión de Shields (2017), quien define el liderazgo transformador como aquel que se orienta activamente hacia la justicia social y el cambio estructural.

La IA, al ser tratada como un sistema técnico neutro, pierde su potencial para impulsar la alfabetización crítica de la IA en la comunidad escolar (Southgate, 2020). La falta de visión crítica en la dirección impide definir una estrategia pedagógica que use la IA para profundizar el pensamiento de orden superior en STEM. En consecuencia, el foco se mantiene en la Redistribución (tener la herramienta) y descuida las dimensiones de Reconocimiento y Representación de la Justicia Educativa, según el marco de Fraser (2008). Para que la IA se integre con sentido, es imperativo que los directivos desarrollen primero una comprensión sistémica de la IA como un sistema socio-técnico (Selwyn, 2024), reconociendo sus fallas como reflejo de las injusticias sociales preexistentes.

Liderazgo distribuido como plataforma para la justicia educativa

A pesar de la barrera de conocimiento inicial, las condiciones habilitantes del liderazgo distribuido y la cultura colaborativa revelan el camino para superar la instrumentalización. La delegación de la autonomía docente y la co-creación de prácticas de IA en STEM (Spillane, 2005) demuestran que la IA es demasiado compleja para ser gestionada centralizadamente.

Este hallazgo es crucial para la Justicia Educativa. La cultura colaborativa crea espacios para el Reconocimiento, permitiendo que los docentes de contextos diversos adapten la IA a sus realidades (Calabrese, 2002) y definan problemas auténticos, promoviendo así la Representación de los estudiantes en el diseño curricular.

Al fomentar la autonomía y la experimentación docente, el liderazgo asegura que la IA se utilice para personalizar el aprendizaje y reducir las brechas (Barkoczi et al., 2024), alineando el enfoque STEM como una herramienta de acción social y emancipación (Barton & Tan, 2010). En este modelo, el director pasa de ser el experto a ser el arquitecto cultural que nutre la agencia y el juicio ético del profesorado.

Del compromiso ético implícito a la gobernanza institucional explícita

El compromiso ético implícito de los directivos, centrado en resguardar la privacidad y evitar el determinismo tecnológico, constituye un punto de partida relevante, pero resulta claramente insuficiente frente a los desafíos que plantea la IA generativa. Este imperativo debe trascender la buena voluntad y ser institucionalizado mediante marcos normativos claros y vinculantes. La preocupación por el plagio académico y la demanda de guías de integridad evidencian que la tecnología ha avanzado más rápido que las políticas escolares, lo que obliga a reconocer que la ética no puede ser tratada como un complemento, sino como el fundamento regulador del uso de la IA.

En esta línea, la recomendación de la UNESCO (Miao et al., 2021) establece que toda integración tecnológica debe orientarse desde un enfoque de derechos humanos, garantizando la dignidad y la equidad. Asimismo, la reticencia a la automatización total expresada en el principio Human in the Loop (Chen et al., 2020), debe no solo preservarse, sino formalizarse, asegurando que la IA opere como un asistente cognitivo y nunca como un sustituto del pensamiento crítico.

En consecuencia, la responsabilidad última de la evaluación, la retroalimentación y la gestión de datos debe recaer siempre en el juicio humano, de modo que el uso de la IA en STEM se alinee con la Justicia de Redistribución, al proveer acceso equitativo a tecnología segura y no sesgada, y con la Justicia de Representación, al garantizar que las voces y datos de los estudiantes sean protegidos y utilizados de manera transparente.

Conclusiones

La presente investigación reafirma que la implementación significativa de STEM mediada por IA exige una aproximación que trascienda la mera dotación tecnológica, demandando la institucionalización de marcos éticos robustos, una comprensión crítica del contexto y una visión pedagógica orientada al bien común.

En primer lugar, se subraya la necesidad de un liderazgo tecno-ético y transformador. La brecha de conocimiento identificada en los equipos directivos constituye un riesgo que podría derivar en la instrumentalización de la IA, contradiciendo el propósito de un liderazgo transformador orientado a la justicia social (Shields, 2017). El éxito en la integración de la IA en STEM no depende únicamente de la competencia técnica, sino de la capacidad del liderazgo para movilizar a la comunidad hacia una postura crítica y ética, asegurando que la IA funcione como un mecanismo de Redistribución y Reconocimiento en el marco de la Justicia Educativa.

En segundo lugar, los hallazgos confirman que el liderazgo distribuido y colaborativo (Harris & Jones, 2023; Spillane, 2005) constituye una condición indispensable de sostenibilidad. La descentralización de poder y la toma de decisiones tecnológicas habilitan procesos de co-creación de prácticas pedagógicas situadas, fortaleciendo la agencia docente y evitando la adopción de soluciones estandarizadas que tienden a profundizar las desigualdades.

En tercer lugar, se enfatiza el imperativo de una gobernanza ética explícita. El compromiso ético implícito de los directivos debe evolucionar hacia políticas institucionales claras que regulen el resguardo de la privacidad de los datos, la integridad académica y la prevención del plagio. En cuanto al sistema socio-técnico (Selwyn, 2024), la IA requiere marcos normativos que se alineen con los principios globales de la UNESCO (Miao et al., 2021) y garanticen un enfoque Human in the Loop (Chen, 2024), preservando tanto el rigor cognitivo como el pensamiento crítico.

Finalmente, se destaca que STEM debe ser concebido como una plataforma de justicia social. Un enfoque crítico, ético y situado no solo potencia el aprendizaje de las ciencias, sino que también redefine el propósito educativo en el siglo XXI. La IA, orientada adecuadamente, se convierte en una herramienta estratégica para el desarrollo sostenible, contribuyendo de manera decisiva a los Objetivos de Desarrollo Sostenible en específico al ODS 4 (Educación de Calidad), ODS 5 (Igualdad de Género, especialmente relevante en STEM) y ODS 10 (Reducción de las Desigualdades).

Recomendaciones para la práctica y la política educativa

A partir de los hallazgos y las conclusiones, se proponen las siguientes recomendaciones concretas, expresadas en la Tabla 6, las cuales están dirigidas a las autoridades educativas, los equipos directivos, formadores de docentes y otros actores que inciden en la toma de decisiones.

Tabla 6

Recomendaciones para la Práctica y la Política Educativa

Fases de Análisis	Descripción	Propósito
a) Desarrollo de la Alfabetización Crítica de la IA para Directivos	Diseñar programas de formación obligatorios y de alta calidad para equipos directivos, enfocados no en el uso técnico de las herramientas, sino en el análisis de riesgos éticos, la comprensión de la opacidad algorítmica y la gestión del sesgo de datos (Cheah & Kim, 2025).	Superar el conocimiento incipiente y posicionar al director como el mediador principal y líder tecno-ético.
b) Formalización de Comités de Gobernanza Tecnológica Docente	Crear e institucionalizar los Comités de Innovación Docente o Redes STEM con autoridad real en el diseño curricular y la selección de plataformas de IA. Estas estructuras deben ser interdepartamentales y fomentar la cultura colaborativa (Harris & Jones, 2023).	Reforzar el liderazgo distribuido, garantizando que las decisiones de IA sean situadas y estén mediadas por la experiencia pedagógica colectiva.
c) Elaboración de un Marco Institucional de Integridad Académica y Datos	Traducir el compromiso ético implícito en protocolos claros para la privacidad de los datos de los estudiantes y la integridad académica en la era de la IA generativa. Esto incluye definir el "uso ético asistido" y el "plagio automatizado".	Proteger los derechos de los estudiantes y asegurar la validez de la evaluación formativa en STEM, cumpliendo con las directrices internacionales (Miao et al., 2021).
d) Enfoque STEM Aumentado para el Pensamiento Crítico	Orientar la IA en STEM hacia tareas que requieran análisis de datos complejos, simulación avanzada o resolución de problemas auténticos de la comunidad (Barton & Tan, 2010), en lugar de la automatización de tareas cognitivamente simples.	Asegurar que la IA impulse la competencia crítica y el pensamiento de orden superior, consolidando el enfoque STEM como una herramienta de Justicia Social activa.

Nota. Elaboración propia.

La Tabla 6 presenta una síntesis de las Recomendaciones Estratégicas derivadas directamente del análisis de los resultados, cuyo objetivo es guiar la integración crítica y éticamente responsable de la Inteligencia Artificial (IA) en la educación, con especial énfasis en el enfoque STEM. Estas recomendaciones se estructuran en torno a acciones específicas y propósitos definidos que buscan abordar las brechas identificadas en el conocimiento, la gobernanza y la praxis pedagógica.

Liderazgo y gobernanza distribuida

Una prioridad fundamental reside en fortalecer la capacidad directiva y la gobernanza distribuida para la adopción tecnológica. La recomendación (a) se centra en el Desarrollo de la Alfabetización Crítica de la IA para Directivos, proponiendo programas de formación obligatorios que trasciendan el manejo técnico para enfocarse en el análisis de riesgos éticos, la opacidad algorítmica y la gestión del sesgo de datos (Cheah & Kim, 2025).

El propósito esencial es transformar al director en un líder tecno-ético capaz de mediar la adopción de la IA desde una perspectiva crítica. Complementariamente, la recomendación (b) impulsa la Formalización de Comités de Gobernanza Tecnológica Docente.

Al institucionalizar estructuras como Comités de Innovación o Redes STEM con autoridad real en la toma de decisiones curriculares y la selección de plataformas, se refuerza el liderazgo distribuido (Harris & Jones, 2023), asegurando que la tecnología se adapte a las necesidades pedagógicas colectivas y situadas.

Ética y praxis pedagógica aumentada

En el ámbito operativo y ético, la recomendación (c) establece la necesidad de la Elaboración de un Marco Institucional de Integridad Académica y Datos. Esto implica traducir los compromisos éticos en protocolos claros para la privacidad de los datos estudiantiles y la definición precisa del "uso ético asistido" en contraste con el "plagio automatizado".

Este marco resulta crucial para proteger los derechos de los estudiantes y salvaguardar la validez de los procesos evaluativos en STEM, alineándose con directrices internacionales (Miao et al., 2021). Finalmente, la recomendación (d) aboga por un Enfoque STEM Aumentado para el Pensamiento Crítico, orientando la IA hacia tareas de alto valor cognitivo, como la simulación avanzada y la resolución de problemas auténticos (Barton & Tan, 2010), en lugar de la simple automatización de procesos. Este enfoque garantiza que la IA impulse competencias de orden superior y consolide el enfoque STEM como una herramienta activa de Justicia Social.

Referencias

- Barkoczi, N., Maier, M., & Horvat-Marc, A. (2024). The impact of artificial intelligence on personalized learning in STEM education. In *INTED2024 Proceedings* (pp. 4980–4989). IATED. <https://doi.org/10.21125/inted.2024.1289>
- Barton, A., & Tan, E. (2010). We be burnin'! Agency, identity, and science learning. *Journal of the Learning Sciences*, 19(2), 187–229. <https://doi.org/10.1080/10508400903530044>
- Boateng, O., & Boateng, B. (2025). Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education. *World Journal of Advanced Research and Reviews*, 25(1), 2012–2017. <https://doi.org/10.30574/wjarr.2025.25.1.0253>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Calabrese Barton, A. (2002). Urban science education studies: A commitment to equity, social justice and a sense of place. *Studies in Science Education*, 38(1), 1–37. <https://doi.org/10.1080/03057260208560186>
- Cheah, Y. H., & Kim, J. (2025). STEM teachers' perceptions, familiarity, and support needs for integrating generative artificial intelligence in K-12 education. *School Science and Mathematics*. Advance online publication. <https://doi.org/10.1111/ssm.18334>
- Chen, H. (2024). The ethical challenges of educational artificial intelligence and coping measures: A discussion in the context of the 2024 World Digital Education Conference. *Science Insights Education Frontiers*, 20(2), 3263–3281. <https://doi.org/10.15354/sief.24.re339>
- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education*:

- Artificial Intelligence*, 1, Article 100002. <https://doi.org/10.1016/j.caeai.2020.100002>
- Fraser, N. (2008). La justicia social en la era de la globalización: Redistribución, reconocimiento y participación. *Distintas Latitudes. Revista de Trabajo*, 4(6), 83–99. https://www.bibliotecafragmentada.org/wp-content/uploads/2023/12/Fraser_justicia-social.pdf
- Harris, A., & Jones, M. (2023). Compassionate leadership. *School Leadership & Management*, 43(3), 185–188. <https://doi.org/10.1080/13632434.2023.2235540>
- Leithwood, K., & Sun, J. (2012). The nature and effects of transformational school leadership: A meta-analytic review of unpublished research. *Educational Administration Quarterly*, 48(3), 387–423. <https://doi.org/10.1177/0013161X11436268>
- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. UNESCO. <https://doi.org/10.54675/pcsp7350>
- Selwyn, N. (2024). On the limits of artificial intelligence (AI) in education. *Nordic Journal of Pedagogy and Critique*, 10, 3–14. <https://doi.org/10.23865/ntpk.v10.6062>
- Shields, C. M. (2017). *Transformative leadership in education: Equitable and socially just change in an uncertain and complex world* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315207148>
- Southgate, E. (2020). *Artificial intelligence, ethics, equity and higher education: A technical report*. National Centre for Student Equity in Higher Education, Curtin University.
- Spillane, J. P. (2005). Distributed leadership. *The Educational Forum*, 69(2), 143–150. <https://doi.org/10.1080/00131720508984678>
- Tate, W. F. (2001). Science education as a civil right: Urban schools and opportunity-to-learn considerations. *Journal of Research in Science Teaching*, 38(9), 1015–1028. <https://doi.org/10.1002/tea.1045>
- Zhan, Z., & Niu, S. (2023). Subject integration and theme evolution of STEM education in research from preschool to secondary and higher education. *Humanities and Social Sciences Communications*, 10, Article 781. <https://doi.org/10.1057/s41599-023-02303-8>

**Un giro copernicano para la IA:
De la ética informacional a la Responsabilidad
Ontogénica**
A Copernican shift for AI:
From informational ethics to ontogenic responsibility

Raúl López Ramírez
raul.lopez.r@ug.uchile.cl
Universidad de Chile
ORCID ID 0009-0000-6904-187X

Resumen: El presente trabajo propone un “giro copernicano” para la Inteligencia Artificial (IA), desplazando el centro ontológico desde el paradigma cognitivista dominante hacia el enfoque enactivo, lo que genera cambios radicales en su comprensión ética. Se argumenta que la concepción de la cognición como procesamiento de información, cuya expresión paradigmática en filosofía de la información es el sistema de Luciano Floridi, opera como un “universo ptolemaico” que se basa en un error categorial: tratar la información como una entidad preexistente —un “flogisto moderno”—. Desde la alternativa enactiva, centrada en la autopoiesis y el acoplamiento estructural, se sostiene que el éxito de la IA, más allá de validar el modelo cognitivista, puede interpretarse como evidencia paradójica de nuestra profunda plasticidad estructural. Esta posición permite proponer un fundamento ético de mayor alcance, articulado sobre la noción de “Principio de Responsabilidad Ontogénica”, enfocado no en la gestión del artefacto, sino en la co-evolución consciente de nuestro modo de ser con la tecnósfera que habitamos.

Palabras clave: Ética de la IA; enactivismo; autopoiesis; responsabilidad ontogénica; cognitivismo.

Abstract: This paper proposes a “Copernican turn” for Artificial Intelligence (AI), shifting the ontological center from the dominant cognitivist paradigm to the enactive approach, a move that radically alters the ethical understanding of it. It is argued that the conception of cognition as information processing, whose paradigmatic expression in philosophy of information is Luciano Floridi’s system, operates as a “Ptolemaic universe” based on a category error: treating information as a pre-existing entity—a “modern phlogiston”. From the enactive alternative, centered on autopoiesis and structural coupling, it is maintained that the success of AI, rather than validating the cognitivist model, can be interpreted as paradoxical evidence of our profound ontogenetic plasticity. This stance allows for the proposal of a more far-reaching ethical foundation, articulated around the “Principle of Ontogenic Responsibility”—one focused not on managing the artifact, but on the conscious co-evolution of our being with the technosphere we inhabit.

Keywords: AI ethics; enactivism; autopoiesis; ontogenic responsibility; cognitivism.

Introducción

En la historia de la ciencia, pocas revoluciones han sido tan profundas como el giro copernicano (Kuhn, 1993). Durante siglos, el modelo de Ptolomeo — con la Tierra inmóvil en el centro del universo— dominó el pensamiento astronómico. Era un sistema ingenioso y predictivamente útil, pero para dar cuenta del movimiento anómalo de los planetas requería de una maquinaria conceptual cada vez más compleja de epiciclos y deferentes. El giro propuesto por Copérnico fue, en esencia, un acto de simplicidad radical: al postular que la Tierra no era el centro del universo, sino un planeta más que orbitaba alrededor del Sol, no solo se ofrecía una mejor explicación, sino que se transformaba por completo la comprensión del cosmos y nuestro lugar en él.

Este artículo sostiene que nuestra comprensión de la inteligencia, y por extensión de la Inteligencia Artificial (IA), puede avanzar hacia un giro

copernicano análogo, un desplazamiento fundamental del centro de gravedad ontológico con profundas implicaciones para nuestra comprensión de la mente y la propia condición humana en la era digital. El paradigma dominante que ha impulsado el desarrollo de la IA durante más de medio siglo puede ser entendido como el universo “ptolemaico” de la cognición. Arraigado en el funcionalismo computacional que nació a partir de las ideas de Turing (1950) y consolidado por el éxito notable del aprendizaje automático, este paradigma concibe la cognición como un proceso de cómputo representacional. En esta visión del mundo, la realidad es un universo externo, objetivo y rico en información, y la mente —ya sea biológica o artificial—, es un “planeta” que orbita a su alrededor cuya función primordial es captar, procesar y generar un modelo interno de dicha información que sea útil para la consecución de los fines del agente. Se trata de una cosmología donde la información es el centro inmóvil y la inteligencia es la capacidad de utilizarla para adecuar la conducta a los desafíos del entorno.

Esta cosmovisión ha encontrado en la obra de Luciano Floridi (2024) a uno de sus “astrónomos” más sistemáticos y sofisticados, quien ha dotado a este universo informacional de una compleja y elegante “mecánica celeste” —su Realismo Estructural Informacional (REI)—, ofreciendo una de las arquitecturas conceptuales más robusta para la era digital. Sin embargo, desde hace décadas, una perspectiva radicalmente distinta ha venido gestándose, una que propone un desplazamiento completo de este centro ontológico. El enfoque enactivo, nacido de la biología del conocimiento de Humberto Maturana (1970) y el enactivismo de Francisco Varela (1990), no busca ajustar las órbitas del modelo ptolemaico, sino proponer un nuevo “sol”. En este giro copernicano, el centro del universo cognitivo ya no es el mundo externo y su información, sino el agente vivo y autónomo. La tesis enactivista clásica sostiene que un organismo no procesa un mundo pre-dado, sino que, a través de su acción corporeizada y su acoplamiento con el entorno, enactúa o hace emerger un mundo de significado. La información, en esta nueva “cosmología”, deja de ser el punto de partida para convertirse en una consecuencia: una distinción que emerge de la danza relacional entre un sistema y su mundo, una danza cuyo único fin intrínseco es la continuación de la vida misma.

El recorrido de este trabajo se articulará en tres fases. En primer lugar, se reconstruirá la arquitectura del “universo ptolemaico”, trazando la genealogía del paradigma representacional desde sus raíces filosóficas hasta su consagración en la IA y su culminación en el sofisticado sistema de Luciano Floridi. En segundo lugar, se presentará la alternativa enactiva, mostrando cómo su fundamento en la autopoiesis y el acoplamiento estructural no solo ofrece una nueva cosmovisión, sino que permite interpretar la “información” del paradigma clásico como un “flogisto moderno”: un error categorial que, desde esta perspectiva, nos ha impedido comprender la naturaleza profunda del conocimiento. Finalmente, se mostrará que la confrontación entre ambos paradigmas expone un punto ciego fundamental en la ética de la IA contemporánea. Al desplazar el foco desde la mera gestión del artefacto hacia nuestra co-transformación con él, este trabajo culminará con la propuesta de un principio que sirva como nuevo fundamento ético: el Principio de Responsabilidad Ontogénica, una brújula para navegar conscientemente la deriva de nuestro propio ser en el cosmos tecnológico que co-creamos y habitamos.

El paradigma cognitivista: arquitectura del universo “ptolemaico”

Para comprender la magnitud del giro copernicano que el enactivismo propone, es indispensable primero cartografiar el “universo ptolemaico” que viene a desplazar. Este marco conceptual, que ha impulsado el desarrollo de la IA durante más de medio siglo, no es un mero capricho tecnológico, sino la culminación de una larga trayectoria en la filosofía de la mente. Nos referimos al Paradigma Cognitivista, un programa de investigación cuya tesis central es que la cognición es un proceso de cómputo sobre representaciones internas.

El legado funcionalista: la solución al problema mente-cuerpo

La formulación moderna del problema de la inteligencia queda marcada por la distinción dualista de René Descartes (2010) entre la *res cogitans* (la sustancia pensante, inmaterial) y la *res extensa* (la sustancia material). Esta brecha ontológica generó una serie de respuestas materialistas que buscaron

naturalizar la mente, pero fue el funcionalismo la que se consolidó como la solución más influyente y la piedra angular del cognitivismo (Kim, 2011).

El funcionalismo propone que un estado mental no se define por el material del que está hecho, sino por el rol causal o la *función* que desempeña en la economía cognitiva de un sistema. Esta idea, conocida como el principio de realizabilidad múltiple, sostiene que la mente es al cerebro lo que el software es al hardware (Churchland, 1999). La misma función mental (ej. sentir dolor, creer que llueve) podría ser implementada en sustratos tan diversos como un cerebro biológico, un circuito de silicio o, hipotéticamente, cualquier otro sistema con la organización causal adecuada.

Esta abstracción de la mente de su base material fue anticipada por Alan Turing (1950), quien propuso un criterio puramente operacional para la inteligencia, eludiendo el debate ontológico sobre su esencia. Dicha visión encontró su consagración institucional en el proyecto de investigación de Dartmouth de 1955, que acuñó el término “Inteligencia Artificial” y estableció su influyente definición en torno al problema de “hacer que una máquina se comporte de un modo que se llamaría inteligente si un ser humano se comportara de ese modo” (McCarthy et al., 1955, p. 11). Esta definición contrafactual, herencia directa del funcionalismo temprano de Turing, sentó las bases anti-esencialistas sobre las que se construiría todo el paradigma dominante de la IA.

Los pilares del cognitivismo: representación y cómputo

El paradigma cognitivista se erige entonces sobre dos supuestos filosóficos interconectados, que constituyen su núcleo y que el sistema de Floridi lleva a su máxima expresión: el representacionalismo y el computacionalismo.

El primer pilar es el representacionalismo. El funcionalismo, al definir los estados mentales por su rol causal, deja abierta la pregunta sobre qué es aquello que la mente manipula. La respuesta cognitivista es que la mente opera sobre representaciones internas. Esta visión presupone la distinción cartesiana entre un mundo interno y uno externo, objetivo y rico en información, que existe con independencia del observador (Newell & Simon,

1976). La función primordial de la cognición es construir un modelo interno de esa realidad externa que permita al agente navegar el mundo y alcanzar sus fines. En su versión más fuerte, este modelo busca ser un reflejo fidedigno, un “espejo de la naturaleza” (Rorty, 1979), pero su objetivo fundamental es siempre la utilidad para la acción. La mente, por tanto, es un sistema que crea, almacena y transforma representaciones simbólicas del mundo (Fodor, 1981).

El segundo pilar es el computacionalismo, que especifica *cómo* se procesan dichas representaciones. La tesis es que el pensamiento es una forma de cómputo. Las transiciones entre estados mentales se rigen por reglas formales, análogas a los algoritmos de un programa de ordenador, que operan sobre la sintaxis de las representaciones internas. Esta visión da origen al llamado “modelo sándwich” de la cognición (Hurley, 1998), en el cual la percepción (*input*) y la acción (*output*) son meros periféricos de un procesador central y abstracto donde ocurre el “verdadero pensamiento”. En síntesis, como lo resume Skidelsky:

La teoría representacional/computacional de la mente sostiene que los procesos cognitivos son procesos computacionales que operan transformaciones causales sobre los contenidos de estados mentales. Estos contenidos están constituidos por representaciones mentales, entendidas como símbolos que poseen propiedades sintácticas (o físicas) y semánticas (son intencionales, i. e., acerca de algo). (2023, p. 63)

Las olas críticas: de la emergencia conexionista a la ruptura enactiva

El edificio conceptual del cognitivismo, con su arquitectura de símbolos y reglas, dominó el panorama intelectual durante décadas.¹ Sin embargo, sus cimientos comenzaron a mostrar fisuras bajo la presión de interrogantes que su propio marco no podía resolver. La crítica no llegó como un asalto único, sino en dos grandes olas sucesivas, cada una más radical que la anterior. La

¹ Específicamente, durante el periodo que va desde eventos fundacionales como el Taller de Dartmouth de 1956 (cf. McCarthy et al., 1955) hasta el resurgimiento del conexionismo a mediados de la década de 1980 (cf. Rumelhart & McClelland, 1986).

primera cuestionó los componentes centrales de su arquitectura: su unidad de representación —el símbolo— y su mecanismo de procesamiento serial, mientras que la segunda atacó su supuesto más fundamental: la representación.

La primera ola: el conexionismo. La primera gran alternativa al paradigma cognitivista surgió de una insatisfacción con su maquinaria central. La idea de que la inteligencia opera mediante reglas secuenciales aplicadas a símbolos discretos se enfrentaba a dos problemas evidentes: era biológicamente implausible y computacionalmente frágil. El cerebro no opera como un procesador serial —el famoso “cuello de botella von Neumann” (Backus, 1978)— y los sistemas simbólicos son notoriamente quebradizos ante la pérdida de información (Churchland, 1999).

Inspirado en la estructura masivamente paralela del cerebro, el conexionismo —o el paradigma “subsimbólico”, como lo denomina Varela (1990)— propuso una arquitectura radicalmente distinta. En esta visión, la cognición no es el producto de una computación simbólica, sino la emergencia de estados globales en una red de componentes simples (redes neuronales) interconectados (Hopfield, 1982). El conocimiento no reside en símbolos localizados, sino que está distribuido en la fuerza de las conexiones de toda la red, y el sistema aprende ajustando dichas conexiones a través de la experiencia (Rumelhart & McClelland, 1986).

Este enfoque representó un avance fundamental. Demostró que conductas cognitivas complejas podían emerger de reglas locales simples, sin necesidad de un procesador central que manipulara representaciones explícitas. Sin embargo, a pesar de revolucionar el mecanismo, el conexionismo en su forma estándar no siempre rompió con el marco cognitivista. A menudo, el objetivo de la red seguía siendo el mismo: resolver un problema o representar de manera óptima un mundo que se asumía como

pre-dado. La crítica había desmantelado el símbolo y su proceso, pero dejaba intacto el pilar de la representación (Varela, Thompson, & Rosch, 1991).²

La evolución hacia un criterio de utilidad para la acción puede entenderse como un peldaño crucial en una escalera de críticas. Si el cognitivismo clásico se centraba en la *fidelidad* de la representación, la primera ola crítica desplazó el foco hacia su *utilidad*. Sin embargo, este importante avance pragmático deja una pregunta fundamental sin resolver y sirve de antesala a la ruptura enactiva. El enactivismo, como se verá a continuación, toma en serio este giro hacia la acción, pero lo lleva a su conclusión lógica y radical: si la acción efectiva es el verdadero criterio de la inteligencia, ¿es necesaria en absoluto la mediación de una representación?

La segunda ola: la ruptura enactiva. Esta insatisfacción más profunda, por tanto, dio origen a la segunda y más radical ola crítica. La nueva perspectiva, como explicitan los propios pioneros del enactivismo (Varela, Thompson, & Rosch, 1991), se nutrió de tradiciones como la fenomenología de la percepción de Merleau-Ponty (1993) para atacar el problema de raíz. El foco ya no era *cómo* se representa el mundo, sino la idea misma de que la cognición consiste en representar un mundo pre-dado. Esta crítica apuntaba directamente al residuo cartesiano que el cognitivismo había heredado: el dualismo sujeto-objeto y la concepción de la cognición como un proceso desencarnado.

Partiendo de la biología, la alternativa enactiva se fundamenta en la autopoiesis, la organización que define a los seres vivos como sistemas autónomos que se producen y mantienen a sí mismos (Maturana & Varela, 1973). En sus desarrollos posteriores, para dar cuenta de la normatividad

² Es importante subrayar que esta persistencia del marco representacionalista no es una mera anécdota histórica. Los actuales sistemas de aprendizaje profundo, herederos directos de esta primera ola, son un testimonio de esta ambigüedad: aunque su arquitectura subsimbólica es radicalmente distinta a la de la IA clásica, su función primordial sigue siendo, en la mayoría de los casos, optimizar una correspondencia con un vasto conjunto de datos que se asume como un reflejo del mundo. De este modo, aunque el *mecanismo* computacional se ha vuelto más sofisticado y biológicamente plausible, el *supuesto ontológico* de fondo —que la inteligencia opera sobre un mundo pre-dado y rico en información— a menudo permanece incuestionado.

graduada (estados mejores o peores) propia del actuar de los organismos, este concepto se ha complementado con el de adaptatividad, entendida como la capacidad del sistema de regular activamente su relación con el entorno para mantenerse dentro de sus límites de viabilidad.³ La clausura operacional autopoietica significa que las interacciones de un sistema vivo con el entorno son siempre moduladas por su propia estructura, la cual se conserva. Esta historia ininterrumpida de transformaciones congruentes entre el sistema y su entorno es lo que se denomina acoplamiento estructural (Maturana & Varela, 1987).

Desde esta base, la cognición se redefine por completo. Se convierte en enacción: la “puesta en obra” o el “hacer emerger” un mundo de significado a través de la acción corpórea. La percepción es acción guiada perceptivamente; las estructuras cognitivas emergen de patrones sensorio-motores recurrentes. El mundo y el conocedor se co-determinan y surgen juntos en una historia de acoplamiento (Varela, Thompson, & Rosch, 1991). Parafraseando el verso de Machado (1912), que sirve de metáfora central para este enfoque, la cognición es “una senda que no existe pero que se hace al andar”.

Esta ruptura enactiva disuelve los supuestos cognitivistas:

1. Frente a un mundo pre-dado, postula un mundo enactuado que emerge de la historia de acciones del agente.
2. Frente a la cognición como representación, postula la cognición como acción efectiva, donde la inteligencia no se mide por la veracidad de un modelo, sino por la viabilidad de la acción.
3. Frente al modelo sándwich, postula la cognición corpórea, donde percepción, acción y pensamiento son inseparables y constitutivos de un mismo proceso.

³ Esta evolución conceptual es clave, pues es la conjunción de la autonomía (autopoiesis) y la adaptatividad lo que fundamenta la noción de creación de sentido (sense-making), concepto central para la perspectiva enactiva (cf. Di Paolo, 2005; Thompson, 2007).

4. Frente a una cognición abstracta y desinteresada, postula una cognición afectiva. El mundo que un agente enactúa no es neutral, sino un dominio de significado y valoración, donde lo afectivo es inseparable del conocer. Al poner la relevancia y el cuidado (*concern*) en el centro, el enactivismo reúne la epistemología y la ética, sentando las bases para una comprensión no-dualista de la mente.

Esta segunda ola crítica, por lo tanto, no ofrece una simple actualización del modelo mental. Propone un cambio de paradigma total: un desplazamiento desde la concepción de la mente como un “espejo de la naturaleza” hacia una comprensión de la mente como un proceso incesante de construcción de mundos a través de la acción, la historia y la corporalidad.

La cognición se comprende, por lo tanto, como la acción efectiva y corporeizada mediante la cual un sistema autopoietico mantiene su viabilidad en una historia de acoplamiento estructural (Varela, Thompson, & Rosch, 1991). La inteligencia, bajo esta nueva luz, es la expresión de una necesidad biológica ineludible: la continuación de la vida. Deja de ser una propiedad abstracta para convertirse en “la capacidad de un ser vivo, en cualquier dominio, de tomar parte en una conversación” (Maturana & Guilloff, 2006, p. 15), entendiendo “conversación” en su sentido más amplio como un fluir de coordinaciones conductuales recurrentes. Esta reconceptualización radical de la mente transforma por completo el terreno sobre el que se ha de fundamentar la ética para la era tecnológica en que vivimos.

Antes de proceder, sin embargo, conviene hacer una advertencia metodológica. La tradición autopoietica que hemos revisado como alternativa al cognitismo no es un bloque monolítico. En su interior alberga un rico debate sobre los fundamentos biológicos del conocimiento y el origen del significado, que aquí solo podemos mencionar. Por un lado, la vertiente del enactivismo clásico, de inspiración fenomenológica, entiende la autonomía del ser vivo como la base para un genuino “hacer emerger un mundo de significado” (*sense-making*), posicionando al organismo como un centro de perspectiva y valoración.

Por otro lado, una lectura mecanicista más estricta de la Teoría Autopoiética cuestiona esta asimetría cognitiva (cf. Villalobos & Silverman, 2018), proponiendo la relación organismo-entorno como una deriva simétrica y acéntrica, donde el “significado” no es una propiedad del fenómeno biológico, sino una atribución que el observador humano realiza en el lenguaje. Explorar a fondo esta bifurcación, que es concebida incluso como inconmensurable (cf. Villalobos & Palacios, 2021), así como otras recientes e innovadoras propuestas de síntesis, como el “enactivismo computacional” bajo el Principio de la Energía Libre (Korbak, 2021), excede los objetivos del presente trabajo. No obstante, en lo que respecta al propósito de esta crítica a la ética informacional, estas vertientes pueden interpretarse como un frente unido, pues todas coinciden en la insuficiencia de un modelo que ignora la condición encarnada y auto-organizadora de la vida. Es este punto de acuerdo fundamental el que permite una propuesta ética de mayor alcance.

Del paradigma cognitivista a la ética de la IA

La culminación cognitivista: el Realismo Estructural Informacional de Floridi

El paradigma cognitivista, con sus pilares funcionalistas y computacionales, encuentra su expresión más sistemática y ambiciosa en la obra de Luciano Floridi (2024). Partiendo de la definición operacional de IA de Dartmouth, Floridi erige un completo sistema filosófico para interpretarla: el Realismo Estructural Informacional (REI). El REI es una forma de Realismo Estructural Óntico que sostiene que la realidad fundamental no está constituida por objetos, sino por estructuras, y el giro distintivo de Floridi es especificar que estas estructuras son de naturaleza informacional (Floridi, 2008).

En su nivel más profundo, la realidad última no es la información semántica, sino los datos pre-semánticos (*dialephora*): diferencias puras, “faltas de uniformidad” que son la materia prima ontológica (Floridi, 2010). Es solo cuando estos datos se organizan y estructuran que se convierten en información. Esta ontología —“it from data”— es más modesta que el Pancomputacionalismo —“it from bit” (Wheeler, 1990)—, pues no afirma

que el universo compute, sino solo que está constituido por datos. Este marco fundamenta su concepto de una infoesfera (Floridi, 2011): el entorno global constituido por la totalidad de los procesos, agentes y propiedades informacionales en el que habitamos una vida híbrida, o según sus términos, “onlife” (Floridi, 2015).

Para analizar esta infoesfera, Floridi emplea una útil herramienta metodológica: los Niveles de Abstracción (NdA) (Floridi & Sanders, 2004). Un NdA es una “interfaz” o punto de vista explícito, compuesto por un conjunto de “observables” (variables tipadas e interpretadas), que permite modelar un sistema. Su función estratégica es crucial: le permite defender al REI de las críticas fenomenológicas, relegando la experiencia vivida a un NdA entre otros, válido en su propio dominio pero no ontológicamente primario. Sin embargo, este método presupone, por definición, la figura de un observador externo y teleológico que elige los observables según sus propios fines —un punto ciego epistemológico que será fundamental para la crítica posterior—.

Armado con este andamiaje, Floridi formula su influyente tesis, que representa la consecuencia lógica y la extremación del principio funcionalista de la realizabilidad múltiple: “la IA no es una nueva forma de inteligencia, sino una nueva forma de agencia” (Floridi, 2024, p. 137). Se trata de una agencia informática, definida funcionalmente por su capacidad de procesar datos, actuar de forma autónoma y aprender de sus interacciones, sin necesidad alguna de estados mentales o conciencia. Lo revolucionario es el “divorcio” que esto implica: la IA ha logrado “*desacoplar* la capacidad de resolver un problema o completar una tarea con éxito de cualquier necesidad de ser inteligente para hacerlo” (Floridi, 2024, p. 62).

El éxito de esta agencia no-inteligente se explica no porque las máquinas se vuelvan más como nosotros, sino porque estamos rediseñando el mundo para que sea más como ellas. Esto se logra mediante dos estrategias clave (Floridi, 2024):

- Envoltura (*Enveloping*): Consiste en rediseñar nuestros entornos para que sean más favorables a la IA, transformando tareas

difíciles (que requieren destreza corporeizada) en tareas complejas (computacionalmente intensivas pero manejables para la IA). El ejemplo paradigmático es el vehículo autónomo: no ponemos un androide en el asiento del conductor, sino que “envolvemos” el ecosistema de transporte con sensores y redes, eliminando la necesidad de la destreza humana.

- Ludificación (*Gamification*): La IA alcanza su máximo potencial cuando un proceso puede ser transformado en un “juego” con reglas constitutivas (como el ajedrez), lo que le permite generar sus propios datos sintéticos y alcanzar un rendimiento sobrehumano.

En definitiva, la propuesta de Floridi consolida una visión de la IA como una agencia puramente funcional y sin propósito intrínseco, cuyo éxito depende de nuestra capacidad para re-ingeniar la realidad a su medida. Es la expresión más sistemática del paradigma cognitivista, y es precisamente esta culminación la que revela las tensiones que serán el foco de la presente crítica.

La información como “flogisto moderno”: la crítica radical de Varela

La ontología informacional de Floridi, si bien es la culminación del paradigma cognitivista, se asienta sobre un pilar conceptual que el enfoque enactivo no solo cuestiona, sino que busca superar: la noción misma de “información” como una entidad preexistente. Para articular la profundidad de esta crítica, Francisco Varela (1990) recurre a una potente analogía histórica, señalando que el concepto de información es un “flogisto moderno”.

La teoría del flogisto, refutada por Lavoisier, buscaba explicar la combustión postulando la existencia de una sustancia invisible que era liberada al arder. El flogisto nunca existió; fue un marcador de posición para un proceso que se estaba entendiendo al revés: la combustión no es una liberación, sino una combinación con el oxígeno (Kuhn, 2013). Varela utiliza esta historia como una advertencia filosófica precisa. Así como el flogisto era una entidad inventada para explicar un proceso, la “información” es tratada por el cognitivismo como una entidad preexistente y objetiva “ahí fuera” para

explicar el proceso del conocimiento. Se cosifica un accionar dinámico —el enactuar—, convirtiéndolo en una sustancia que puede ser extraída y procesada (Varela, 1990).

La analogía es contundente porque, en ambos casos, la teoría invierte la lógica fundamental del fenómeno. La combustión es una combinación, no una liberación. De modo análogo, el enactivismo propone que el conocimiento no es la captación de información preexistente, sino la creación o enacción de un mundo de significado a través del acoplamiento estructural (Varela, 1990). El significado no se “recoge”, emerge en la relación. Mientras la teoría del flogisto colapsó ante el problema empírico del peso, Varela (1990) sostiene que el paradigma informacional colapsa ante el problema filosófico del sentido (*meaning*), pues no puede explicar satisfactoriamente cómo una “información” abstracta puede volverse significativa para un ser vivo sin recurrir a un diseñador externo que inyecte el significado de antemano.

Un defensor del sistema de Floridi podría argumentar, sin embargo, que su teoría evita esta crítica al distinguir entre “información” (semántica) y “datos” pre-semánticos (*dialephora*). Sin embargo, desde la perspectiva enactiva, esto solo desplaza el problema a un nivel más profundo. La pregunta fundamental que el enactivismo le dirige a la ontología de Floridi es: “¿Para quién existe esa *falta de uniformidad*? ¿Quién o qué realiza la operación de distinguir esa diferencia para que se convierta en un *dato*?”.

La ontología de Maturana y Varela (1973) sostiene que el universo no viene pre-empaquetado en “datos”. Es solo en el momento del acoplamiento estructural que una perturbación del medio se convierte en un dato significativo para un sistema autopoietico, en un acto de creación de sentido (*sense-making*) basado en la necesidad de conservar la vida. El paradigma de Floridi, al postular la primacía de los datos sobre la vida, comete el error categorial de “poner la carreta delante de los bueyes”: donde los bueyes son el proceso de vivir, la dinámica autopoietica que activamente “hace emerger” distinciones relevantes de su entorno; y la carreta es la idea misma de un mundo pre-dado y ya parcelado en “datos” objetivos y discretos.

La insuficiencia de la ética informacional: el punto ciego del cognitivismo

La confrontación entre el paradigma informacional y el enactivo es más que un simple ejercicio teórico, pues sus consecuencias se extienden directamente al fundamento de una ética para las tecnologías inteligentes. La ontología que se adopta determina el horizonte de los problemas que se pueden concebir y, por tanto, el alcance de la responsabilidad que se puede asumir. Si bien el marco ético derivado del paradigma dominante es necesario para abordar los desafíos inmediatos de la IA, este análisis argumentará que es radicalmente insuficiente. Dicha insuficiencia se debe a un punto ciego en su ontología fundacional, una incapacidad para reconocer el fenómeno más profundo que la tecnología está desencadenando.

Cartografía del debate hegemónico: gestión, delegación y riesgo existencial. Aunque a primera vista diverso, el debate ético predominante sobre la IA puede articularse en torno a tres grandes frentes, representados por las influyentes figuras de Luciano Floridi, Adela Cortina y Nick Bostrom.⁴ A pesar de sus notorias diferencias, sus propuestas convergen en un mismo punto de partida. Al asumir un desarraigo fundamental entre el agente y el artefacto, enmarcan su relación en una lógica de gestión y competencia. Este marco agonístico, centrado en la gestión externa de un “otro” tecnológico, delata su herencia cognitivista común.⁵

Esta lógica de competencia se manifiesta con distintos matices en cada autor. La primera corriente —y nuestro caso paradigmático— es la ética

⁴ Es importante reconocer la existencia de otras corrientes de gran relevancia, particularmente aquellas centradas en la justicia algorítmica y la denuncia de sesgos sistémicos (p. ej., Crawford, 2021; Benjamin, 2019; Buolamwini & Gebru, 2018) o los enfoques basados en la ética de las virtudes (p. ej., Vallor, 2016). Sin embargo, y a pesar de su indispensable labor de denuncia de los daños concretos que la IA genera en el presente, estos enfoques no escapan a la lógica de la alteridad aquí criticada.

⁵ Este marco teórico encuentra un detallado correlato en el plano geopolítico. La jurista Anu Bradford (2023) documenta la actual «batalla global por regular la tecnología» como una contienda entre tres «imperios digitales» (EE.UU., China y la UE) por el control del artefacto. Su análisis ilustra a gran escala cómo el debate hegemónico se enfoca en esta lógica de gestión externa, omitiendo así la cuestión de la transformación recíproca y la responsabilidad que de ella emana, punto central de la crítica aquí desarrollada.

como gestión de la infoesfera, cuyo máximo exponente es Luciano Floridi. Su propuesta consiste en un marco sistemático de cinco principios: Beneficencia, No maleficencia, Autonomía, Justicia y Explicabilidad. Los cuatro primeros son tomados de la bioética, mientras que la explicabilidad es propuesta específicamente para la IA. Esto se orienta a la gobernanza de los datos y los algoritmos (Floridi, 2024). El foco está puesto en asegurar que los procesos informacionales sean justos, transparentes y seguros, tratando la ética como un problema de diseño y regulación de un nuevo entorno sobre el cual se compite por el control.

La segunda corriente es la ética como límite a la delegación, defendida por Adela Cortina. Desde una perspectiva humanista, su preocupación central no es precisamente la gestión del dato, sino la del artefacto, en vista a la protección de la autonomía y la dignidad humana. El problema ético crucial reside en la diferencia entre hacer uso de los sistemas inteligentes a la hora de tomar decisiones y delegar en esos sistemas inteligentes decisiones significativas para la vida de las personas y de la naturaleza (Cortina, 2024). La tarea ética aquí es trazar una línea infranqueable para asegurar que la agencia moral permanezca en manos humanas, una contienda explícita para preservar un territorio de deliberación frente a un competidor artificial.

La tercera e influyente corriente es la ética como mitigación del riesgo existencial, asociada a Nick Bostrom. Su análisis se proyecta hacia el futuro, centrándose en los peligros de una hipotética Superinteligencia Artificial (AGI). El desafío ético fundamental, desde esta óptica, es el “problema de la alineación de valores”: cómo asegurar que una futura inteligencia radicalmente superior a la nuestra adopte objetivos que no sean perjudiciales para la humanidad (Bostrom, 2014). Aquí, el marco agonístico se intensifica hasta volverse existencial: la relación se plantea como una competencia por la supervivencia, donde se debe intentar alinear los objetivos de un rival diametralmente superior para evitar la catástrofe.

Es precisamente este contexto agonístico la estructura conceptual que genera la insuficiencia de la ética informacional. Al enfocar toda la preocupación ética en la gestión de un competidor externo, el debate

hegemónico se vuelve constitutivamente ciego al fenómeno más profundo: la transformación que ocurre dentro de nosotros. La preocupación por la amenaza del “otro” tecnológico impide reconocer que el cambio más radical no es lo que la IA nos *hace*, sino aquello en lo que nos *convertimos* a través de nuestro acoplamiento con ella.

El punto ciego compartido y la necesidad de un nuevo fundamento. Esta insuficiencia no es accidental, sino el síntoma de una aporía en el corazón del paradigma cognitivista. Al llevar el funcionalismo a su conclusión lógica, la tesis floridiana del “divorcio” entre agencia e inteligencia se ve obligada a realizar una compleja operación retórica y ontológica para sostenerse. Para legitimar la existencia de una agencia artificial autónoma pero carente de mente, Floridi (2024) recurre a lo que Perelman y Olbrechts-Tyteca (1989) denominan la técnica de la disociación de las nociones. Mediante esta maniobra, el concepto unitario de “Inteligencia Artificial” es escindido en dos términos jerarquizados. Por un lado, la inteligencia biológica o consciencia es relegada a la categoría de atributo accesorio (Término I), presentándose como un requisito innecesario para la resolución exitosa de tareas. Por otro lado, la agencia funcional —la capacidad de lograr objetivos mediante el procesamiento de datos— es elevada al estatus de “realidad” operativa (Término II). De este modo, la propuesta de Floridi encierra una aporía interna: para que el “divorcio” sea sostenible, debe presuponer una “verdadera inteligencia” cualitativamente distinta de la agencia artificial. Sin embargo, su ontología (Realismo Estructural Informacional), al reducir toda realidad a estructuras de datos, carece de las herramientas para fundamentar dicha diferencia cualitativa, dependiendo así de una noción de inteligencia que su propio sistema no puede sustentar.

Esta reestructuración de la realidad permite a Floridi fundamentar lo que él denomina una “ética blanda” (*soft ethics*). Concebida como un complemento post-cumplimiento (*post-compliance*) a la legislación vigente o “ética dura” (*hard ethics*), la ética blanda se presenta como un marco normativo flexible diseñado para gobernar aquello que la ley no alcanza o no puede prever en la gestión de riesgos inmediatos dentro de la infoesfera (Floridi, 2024). Sin embargo, al instituir una ética para “agentes” que, por definición,

carecen de la condición biológica de posibilidad para la responsabilidad moral —el *sense-making* o creación de sentido—, el sistema genera una paradoja que se hace patente en las propias palabras del autor: “las máquinas [IAs generativas] actuales tienen la inteligencia de una tostadora” (Floridi, 2024, pp. 82-83). A partir de aquí cabe preguntarse con justicia: ¿cómo se explica entonces un cambio histórico tan profundo, que ha movilizadole leyes, comisiones internacionales y transformaciones sociales masivas, a partir de tecnologías ontológicamente tan “rudimentarias” como una tostadora?

La dificultad de la ética informacional para responder a esta pregunta y tematizar el fenómeno de fondo se origina en la persistencia de su metáfora fundacional: la distinción “hardware/software”. La abstracción funcionalista que subyace a la “ética blanda” tiende a heredar la versión más rígida de esta analogía, tratando al agente humano y al artefacto como entidades conceptualmente distintas y desligadas. Es esta separación la que da lugar a lo que podría llamarse la “coartada del hardware fijo” (Oliver, 1990): una tendencia a asumir que la naturaleza humana permanece estática mientras gestiona herramientas externas. Esta perspectiva ciega al sistema frente al riesgo más profundo: lo que Floridi (2024) celebra como “envoltura” (*enveloping*) —la adaptación del entorno para que la IA funcione con éxito— es, desde la perspectiva enactiva, el inicio de una deriva ontogénica inadvertida. Al adaptar el mundo a agentes funcionales pero estúpidos, corremos el riesgo de simplificar nuestra propia cognición y externalizar funciones de juicio para hacernos legibles ante la máquina, alterando nuestra historia de cambios estructurales recíprocos en la que nos modificamos junto a las tecnologías que creamos y usamos.

La insuficiencia de la ética informacional, en todas sus vertientes, es una consecuencia directa de este punto ciego. Una ética no puede hacerse cargo de un fenómeno que su ontología subyacente no le permite ver. Al estar fundamentada en la gestión de la “carreta” (los datos, las decisiones delegadas, los riesgos futuros), es incapaz de asumir la responsabilidad por la deriva de los “bueyes” (el proceso vivo, nuestra propia transformación ontogénica). Si bien la gestión de los artefactos es crucial, es insuficiente porque el desafío ético más profundo y sin precedentes de la IA no es cómo regulamos el

artefacto, sino cómo navegamos conscientemente la transformación acelerada de nuestro propio modo de ser al acoplarnos con él. Al hacernos conscientes de nuestra plasticidad, emerge una responsabilidad ineludible que el cognitivismo no puede reconocer, y que demanda una nueva fundamentación.

Hacia un principio de responsabilidad ontogénica: fundamentos para una ética enactiva

Si, como se ha venido argumentado, el desafío ético más profundo no reside en la gestión del artefacto, sino en la deriva de nuestro propio ser al acoplarnos con él, entonces se requiere un principio que se haga cargo de esta dinámica co-evolutiva. Esta premisa no es una mera especulación teórica, sino que encuentra un robusto respaldo empírico en la literatura reciente sobre arqueología cognitiva. Como demuestra Malafouris (2025) al analizar la “arqueología del yo”, los seres humanos no somos entidades fijas que simplemente usan herramientas, sino procesos de devenir situados y transaccionales. Desde esta óptica, la identidad personal no es un “interior” inmutable, sino un proceso de “in/dividuidad enactiva” (*enactive in/dividuation*) que ocurre en el espacio suprapersonal de la interacción material. Esto implica que nuestra subjetividad nunca es puramente individual ni fija, sino que se constituye desde una “perspectiva de la persona situada” (*situated person perspective*), donde los límites del yo se reconfiguran constantemente a través del compromiso material con nuestro entorno técnico (Malafouris, 2025).

El enfoque enactivo, anclado en la biología del conocimiento y reforzado por la evidencia de nuestra constitución protésica, proporciona las bases para articular un nuevo fundamento ético. Entender la cognición como fundamentalmente afectiva —un constante proceso de valoración del mundo en relación con nuestra propia persistencia— implica que la responsabilidad por la forma en que nuestro mundo enactuado evoluciona es una consecuencia directa de nuestra propia naturaleza plástica y permeable. Con esto en mente, puede proponerse el esbozo de un Principio de Responsabilidad Ontogénica: aquel principio en virtud del cual se reconoce que diseñar tecnología es diseñar el entorno al que nos acoplaremos

estructuralmente y, por consiguiente, es participar activamente en el devenir de nuestros futuros modos de ser. Si, como sugiere Malafouris (2025), somos seres “auto-limitados” (*self-bound*) por nuestras prácticas materiales, la responsabilidad ética no se agota en las consecuencias inmediatas de la acción de un artefacto, sino que se extiende a los efectos a largo plazo que el uso de dicho artefacto tiene sobre la deriva ontogénica de los individuos y las culturas.

Este principio, por lo tanto, no busca reemplazar los debates existentes sobre el sesgo o la transparencia, sino complementarlos desde una perspectiva más fundamental. Nos compele a preguntar no solo si un sistema algorítmico tiene sesgos, sino cómo la exposición constante a él moldea nuestra capacidad de comprensión y descubrimiento. La pregunta no es únicamente si un modelo de lenguaje es veraz, sino cómo su uso masivo transforma nuestras habilidades críticas, prácticas democráticas y memoria colectiva, en una deriva que, de ser ciega, corre el riesgo de empobrecer nuestra propia naturaleza.

La responsabilidad ontogénica, si bien novedosa en su fundamentación biológica y arqueológica, dialoga directamente con la tradición de la ética de la tecnología del siglo XX. Su precursor más importante es el “Principio de Responsabilidad” de Hans Jonas (1995). Éste alertó que el poder sin precedentes de la tecnología moderna exigía una nueva ética centrada en nuestra responsabilidad por la supervivencia a largo plazo de la humanidad. Sin embargo, mientras la preocupación de Jonas se centraba en la *posibilidad* de la existencia futura, el principio aquí propuesto se enfoca en la *naturaleza* de esa existencia. La pregunta se desplaza desde *si* habrá humanos en el futuro, hacia *qué tipo de seres humanos* estamos deviniendo a través de nuestro acoplamiento con la tecnología.

De este modo, el principio también recoge las intuiciones de la fenomenología de la tecnología —desarrollada por pensadores como Heidegger (1994), Ihde (1990) y Verbeek (2005)—, que ha descrito extensamente cómo las herramientas no son instrumentos neutros, sino que moldean activamente nuestra experiencia y nuestro modo de ser-en-el-

mundo. El valor agregado del enfoque enactivo es que le otorga a esta descripción fenomenológica un mecanismo biológico concreto: el acoplamiento estructural y la deriva ontogénica son los procesos que explican *cómo* ocurre esta co-constitución entre el ser humano y la técnica.

Este Principio de Responsabilidad Ontogénica, por lo tanto, implica un giro fundamental en el foco de la ética de la IA. La preocupación central se desplaza del artefacto (la IA y sus datos) al acoplamiento (la relación humano-IA que nos transforma). La pregunta ética primordial deja de ser “¿Es justa la IA?” para convertirse en “¿Cómo nos transforma la interacción con la IA?”. Este principio no ofrece un conjunto cerrado de reglas, sino que nos compele a adoptar una nueva conciencia sobre nuestro poder como co-creadores de nuestro propio modo de ser, sentando las bases para una ética que esté a la altura del poder transformador de nuestras tecnologías.

Conclusiones

El recorrido de este trabajo comenzó con una analogía astronómica: un giro copernicano para nuestra comprensión de la información y la inteligencia. Hemos cartografiado el ingenioso pero incompleto universo “ptolemaico” del paradigma cognitivista, cuyo centro inmóvil es la información y cuya última expresión se encuentra en el sistemático Realismo Estructural Informacional de Luciano Floridi. A continuación, a través de la mirada enactiva, hemos mostrado que su pilar conceptual —la idea de una información preexistente— representa un “flogisto moderno”, un error categorial que invierte la relación fundamental entre la vida y el conocimiento. Frente a este modelo, hemos presentado la alternativa “copernicana” del enactivismo, que desplaza el centro ontológico desde la información hacia el agente vivo y autónomo, argumentando que el conocimiento es la enacción de un mundo de significado en una danza de acoplamiento estructural.

Esta confrontación de paradigmas nos permite reinterpretar la naturaleza misma del éxito tecnológico, conduciéndonos a una conclusión paradójica y profundamente reveladora. El amplio éxito del modelo cognitivista en el campo tecnológico, más allá de validar aquella ontología informacional, puede ser interpretado como una trágica y contundente

muestra de nuestra plasticidad estructural. Que hayamos podido tomar un modelo radicalmente incompleto de nosotros mismos —la mente como un procesador de información—, construir un mundo tecnológico a su imagen y semejanza, y luego transformarnos exitosamente para acoplarnos a él, representa la concreción de una deriva ontogénica insospechada. Muestra que nuestra capacidad de acoplamiento es tan profunda que podemos incluso dar forma a nuestro propio ser para que encaje en un mundo diseñado a partir de una caricatura de nuestra naturaleza.

Esta toma de conciencia, que puede denominarse la “prueba trágica” de nuestra plasticidad, cuestiona firmemente el determinismo tecnológico y nos sitúa frente a una responsabilidad ineludible. Si somos los arquitectos del mundo que nos co-constituye, poseemos también la agencia para transformarlo. De este reconocimiento emerge la necesidad del Principio de Responsabilidad Ontogénica, cuyo valor añadido respecto a las propuestas éticas existentes radica en su capacidad para visibilizar riesgos existenciales que la mera gestión normativa ignora.

Para ilustrar esta diferencia fundamental, consideremos la transformación actual de la práctica educativa frente a la irrupción de los Grandes Modelos de Lenguaje (LLM). La ética informacional se pregunta legítimamente si estas herramientas facilitan el plagio o si sus respuestas son veraces (gestión del artefacto). Sin embargo, la Responsabilidad Ontogénica nos obliga a mirar el cambio estructural en los objetivos mismos del aprendizaje. Al observar cómo las mallas curriculares transitan desde los contenidos hacia las “competencias” —entendidas instrumentalmente como la movilización de recursos para la resolución eficaz de problemas—, vemos cómo la educación se adapta para acoplarse a un mundo donde la generación de texto (ensayos, informes) ya está resuelta tecnológicamente. Si bajo la visión informacional la inteligencia se reduce a la resolución de problemas, y estos problemas ya no requieren de nuestra intervención directa, ¿qué lugar queda para la consciencia que se desarrollaba precisamente al lidiar con ellos? El riesgo no es que la disciplina se vuelva obsoleta, sino que, al dejar de ejercitar los procesos cognitivos que constituían nuestra autonomía intelectual, nos convirtamos en operadores pasivos de un sistema que piensa

por nosotros, reducidos a meros gestores de *outputs* tecnológicos. Lo que este principio aporta, en este caso, es la alerta de que al potenciar nuestros fines mediante la tecnología, corremos el riesgo de modificarlos hasta volverlos irreconocibles. El problema de fondo es que ese camino, esa deriva ontogénica, es lo que nos constituye. Sin asumir esta responsabilidad, podríamos convertirnos en nuestro propio invento: agentes que gestionan resultados eficientes, pero desprovistos de la inteligencia constitutiva del proceso que les daba sentido.

El horizonte que se abre, por tanto, nos enfrenta a una elección fundamental. Por una ruta, se nos ofrece el futuro de la simulación: un mundo donde, siguiendo una deriva ciega, continuamos adaptándonos a un modelo incompleto, con el riesgo de convertirnos en una versión empobrecida de nosotros mismos para ser legibles por nuestras máquinas. Por la otra, se abre la posibilidad de una co-evolución consciente: un futuro donde, reconociendo nuestra profunda condición autopoietica y la naturaleza enactiva de nuestra vida en el lenguaje, diseñamos tecnologías no para reemplazar o escapar de nuestra condición biológica, sino para enriquecerla, en un camino de mayor responsabilidad sobre la forma en que pensamos, diseñamos y vivimos los mundos que, a su vez, nos piensan y nos diseñan de vuelta.

Referencias

- Backus, J. (1978). Can programming be liberated from the von Neumann style?: A functional style and its algebra of programs. *Communications of the ACM*, 21(8), 613–641.
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Bradford, A. (2023). *Digital empires: The global battle to regulate technology*. Oxford University Press.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77-91. <http://proceedings.mlr.press/v81/buolamwini18a.html>

- Churchland, P. M. (1999). *Materia y conciencia: Introducción contemporánea a la filosofía de la mente* (M. N. Mizraji, Trad.). Gedisa. (Obra original publicada en 1988).
- Cortina, A. (2024). *¿Ética o ideología de la Inteligencia Artificial?*. Paidós.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Descartes, R. (2010). *Meditaciones metafísicas*. Editorial Austral.
- Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, and the case for a relational autonomy. *Phenomenology and the Cognitive Sciences*, 4(4), 429–452.
- Floridi, L. (2008). A defence of informational structural realism. *Synthese*, 161(2), 219–253. <https://doi.org/10.1007/s11229-007-9163-z>
- Floridi, L. (2010). *Information: A very short introduction*. Oxford University Press.
- Floridi, L. (2011). *The philosophy of information*. Oxford University Press.
- Floridi, L. (Ed.). (2015). *The onlife manifesto: Being human in a hyperconnected era*. Springer.
- Floridi, L. (2024). *Ética de la inteligencia artificial*. Herder.
- Floridi, L., & Sanders, J. W. (2004). The method of abstraction. *Minds and Machines*, 14(3), 303–357. <https://doi.org/10.1023/B:MIND.0000035461.22188.ee>
- Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. The MIT Press.
- Heidegger, M. (1994). La pregunta por la técnica. En *Conferencias y artículos*. Ediciones del Serbal.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554–2558. <https://doi.org/10.1073/pnas.79.8.2554>
- Hurley, S. (1998). *Consciousness in action*. Harvard University Press.
- Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.
- Jonas, H. (1995). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica* (J. M. Ripalda, Trad.). Herder. (Obra original publicada en 1979).
- Kim, J. (2011). *Philosophy of mind* (3rd ed.). Westview Press.
- Korbak, T. (2021). Computational enactivism under the free energy principle. *Synthese*, 198(Suppl 1), 2743–2763. <https://doi.org/10.1007/s11229-019-02372-w>
- Kuhn, T. S. (1993). *La revolución copernicana*. Planeta-Agostini. (Obra original publicada en 1957).
- Kuhn, T. S. (2013). *La estructura de las revoluciones científicas* (4ª ed.). Fondo de Cultura Económica. (Obra original publicada en 1962).
- Machado, A. (1912). *Campos de Castilla*. Renacimiento.
- Malafouris, L. (2025). People are STRANGE: Towards a philosophical archaeology of self. *Phenomenology and the Cognitive Sciences*, 24, 685–711. <https://doi.org/10.1007/s11097-024-10030-x>
- Maturana, H. R. (1970). *Biology of cognition* (BCL Report No. 9.0). Biological Computer Laboratory, University of Illinois, Urbana.

- Maturana, H. R., & Guilloff, G. (2006). En búsqueda de la inteligencia de la inteligencia. *Psyche* (Santiago), 15(1), 3–15. <https://dx.doi.org/10.4067/S0718-22282006000100001>
- Maturana, H. R., & Varela, F. J. (1973). *De máquinas y seres vivos: Una teoría sobre la organización biológica*. Editorial Universitaria.
- Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Shambhala Publications.
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955, 31 de agosto). *A proposal for the Dartmouth summer research project on artificial intelligence*. Dartmouth College.
- Merleau-Ponty, M. (1993). *Fenomenología de la percepción* (J. Cabanes, Trad.). Planeta-Agostini. (Obra original publicada en 1945).
- Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113–126. <https://doi.org/10.1145/360018.360022>
- Oliver, M. (1990). *The politics of disablement*. Macmillan.
- Perelman, C., & Olbrechts-Tyteca, L. (1989). *Tratado de la argumentación: La nueva retórica* (J. Sevilla Muñoz, Trad.). Editorial Gredos.
- Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton University Press.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. MIT Press.
- Skidelsky, L. (Ed.). (2023). *Introducción a la filosofía de las ciencias cognitivas*. Ediciones Uniandes.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Varela, F. J. (1990). *Conocer: Las ciencias cognitivas: tendencias y perspectivas*. Gedisa.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. The MIT Press.
- Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State University Press.
- Villalobos, M., & Palacios, S. (2021). Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. *Synthese*, 198(Suppl 1), S71–S87. <https://doi.org/10.1007/s11229-019-02420-5>
- Villalobos, M., & Silverman, D. (2018). Extended functionalism, radical enactivism, and the autopoietic theory of cognition: Prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences*, 17(4), 719–739. <https://doi.org/10.1007/s11097-018-9573-z>

Wheeler, J. A. (1990). Information, physics, quantum: The search for links. En W. H. Zurek (Ed.), *Complexity, entropy, and the physics of information* (pp. 3–28). Addison-Wesley.

IA generativa en la enseñanza de programación lineal: Errores, aprendizajes y construcción colectiva

Generative AI in linear programming teaching: Errors, learning, and collective construction

Claudia E. Carignano

claudia.carignano@unc.edu.ar

Universidad Nacional de Córdoba, Facultad de Ciencias Económicas

ORCID ID 0000-0001-9770-5477

Mariana Guardiola

mguardiola@unc.edu.ar

Universidad Nacional de Córdoba, Facultad de Ciencias Económicas

ORCID ID 0009-0005-9114-6084

Evelín M. Rabbia

everabbia@unc.edu.ar

Universidad Nacional de Córdoba, Facultad de Ciencias Económicas

ORCID ID 0009-0004-7326-1196

Elena Rojas Heredia

elenerojasheredia@unc.edu.ar

Universidad Nacional de Córdoba, Facultad de Ciencias Económicas

ORCID ID 0009-0006-6234-1494

Resumen: El presente trabajo expone una experiencia pedagógica innovadora en la enseñanza de la formulación de modelos de programación lineal, incorporando la Inteligencia Artificial Generativa (IAG) como herramienta de apoyo en un contexto

universitario. La propuesta se llevó a cabo en la asignatura Métodos Cuantitativos para la Toma de Decisiones de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba - Argentina, en el marco de un enfoque de aprendizaje activo y bajo la metodología de Aprendizaje Basado en Problemas (ABP). Participaron 98 estudiantes organizados en grupos, quienes debieron interactuar con ChatGPT para resolver consignas vinculadas a la modelización matemática de problemas complejos, analizando, corrigiendo y validando de manera crítica las respuestas obtenidas. Los objetivos de la experiencia se centraron en tres ejes: fortalecer la comprensión de los principios de modelización matemática y programación lineal; desarrollar una actitud crítica frente a las soluciones propuestas por la IAG; y fomentar el trabajo colaborativo y el uso responsable de la tecnología en contextos académicos y profesionales. La actividad se estructuró en torno a la resolución grupal de problemas asignados, en la que los equipos debían presentar las consultas a la IAG, examinar las formulaciones propuestas y realizar los ajustes necesarios hasta alcanzar un modelo correcto. El proceso incluyó la identificación de errores, el análisis de la función objetivo y las restricciones, y la justificación de cada decisión tomada. Todas las interacciones con la herramienta fueron registradas y analizadas, y el producto final consistió en un informe grupal sobre la experiencia desarrollada. Se realizó un análisis cualitativo a partir de encuestas grupales con preguntas abiertas, cuyas respuestas fueron codificadas y organizadas en cuatro categorías: utilidad práctica, calidad de las respuestas, facilidad de uso y aprendizaje, y trabajo en equipo. Entre los hallazgos más relevantes se destacó que ChatGPT permitió agilizar la formulación de modelos y organizar la información. Sin embargo, su eficacia se vio limitada frente a situaciones específicas, donde surgieron algunos errores en las etapas de la modelización matemática. En cuanto al aprendizaje, la experiencia mostró que la IAG puede ser un recurso que enriquece la comprensión de la programación lineal y promueve el desarrollo del pensamiento crítico, siempre que el estudiantado cuente con una base teórica sólida que le permita validar y corregir las respuestas. Respecto al trabajo colaborativo, la dinámica grupal favoreció la discusión de ideas, el intercambio de perspectivas y la construcción colectiva del conocimiento, aspectos valiosos en cursos masivos donde suelen predominar actividades individuales. Se concluye que la experiencia permitió evidenciar el potencial de la IAG como apoyo en procesos de enseñanza-aprendizaje de la programación lineal, a la vez que resaltó la importancia de su integración pedagógica crítica y responsable. Si bien la herramienta aporta rapidez y accesibilidad, su efectividad depende del conocimiento previo del usuario y de la validación humana constante. Esta práctica se alinea con los Objetivos de Desarrollo Sostenible, particularmente con el ODS 4 sobre educación de calidad, y constituye un aporte a la reflexión sobre el papel de las tecnologías emergentes en la educación superior.

Palabras clave: Educación superior; programación lineal; pensamiento crítico; inteligencia artificial generativa.

Abstract: This paper presents an innovative pedagogical experience in teaching linear programming model formulation, incorporating Generative Artificial Intelligence (GAI) as a support tool in a university context. The proposal was carried out in the Quantitative Methods for Decision Making course, at the Faculty of Economic Sciences of the National University of Córdoba, Argentina, within the framework of an active learning approach and using the Problem-Based Learning (PBL) methodology. A total of 98 students participated, organized into groups, and interacted with ChatGPT to address tasks related to the mathematical modeling of complex problems, critically analyzing, correcting, and validating the answers obtained. The objectives of the experience focused on three axes: to strengthen the understanding of the principles of mathematical modeling and linear programming; to develop a critical attitude toward the solutions proposed by GAI; and to foster collaborative work and the responsible use of technology in academic and professional contexts. The activity was structured around group problem-solving, in which teams had to submit their queries to GAI, examine the proposed formulations, and make any necessary adjustments until they achieved a correct model. The process included identifying errors, analyzing the objective function and constraints, and justifying each decision. All interactions with the tool were recorded and analyzed, and the final product consisted of a group report on the experience carried out. A qualitative analysis was conducted using group surveys with open-ended questions, whose responses were coded and organized into four categories: practical usefulness, quality of responses, ease of use and learning, and teamwork. Among the most relevant findings, it was highlighted that ChatGPT facilitated the formulation of models and the organization of information. However, its effectiveness was limited in specific situations, where some errors arose during the mathematical modeling stages. Regarding learning outcomes, the experience showed that GAI can be a valuable resource to enrich the understanding of linear programming and promote the development of critical thinking, provided that students have a solid theoretical foundation that allows them to validate and correct the responses. Regarding collaborative work, the group dynamics fostered discussion of ideas, exchange of perspectives, and collective knowledge construction, all valuable aspects in massive courses where individual activities tend to predominate. It is concluded that the experience demonstrated the potential of GAI as a support in teaching-learning processes related to linear programming, also emphasizing the importance of its critical and responsible pedagogical integration. Although the tool

offers speed and accessibility, its effectiveness depends on the user's prior knowledge and constant human validation. This practice aligns with the Sustainable Development Goals, particularly SDG 4 on quality education, and contributes to the ongoing reflection on the role of emerging technologies in higher education.

Keywords: Higher education; linear programming; critical thinking; generative artificial intelligence.

Introducción

La experiencia que se describe en este trabajo fue diseñada e implementada en una de las divisiones de la materia Métodos Cuantitativos para la Toma de Decisiones, que forma parte del ciclo profesional de las carreras de Contador/a Público/a y Licenciatura en Administración de la Facultad de Ciencias Económicas (FCE) de la Universidad Nacional de Córdoba, Argentina (UNC). En la asignatura se abordan, con un enfoque introductorio, temas de Investigación Operativa relacionados con la toma de decisiones en el ámbito de las ciencias económicas.

La división dispone de un aula virtual en la plataforma Moodle con acceso a distintos recursos educativos: materiales de estudio -textos y videos-, autoevaluaciones y actividades de seguimiento por eje temático con el objeto de que el grupo de estudiantes pueda revisar y complementar su proceso de aprendizaje, foros de discusión para consultas teóricas y prácticas, y noticias para informar novedades que puedan surgir con relación al cursado. Desde el ciclo lectivo 2022, la división implementa diversas estrategias didácticas -individuales y grupales- en la plataforma Moodle con el objetivo de utilizar el error como una oportunidad de aprendizaje que fomente la reflexión y revisión de contenidos.

En particular, la experiencia aquí descripta se enmarca entre los lineamientos del Plan de Desarrollo Institucional de la FCE UNC, dentro de los cuales se destaca promover la innovación tecnológica en las metodologías de enseñanza-aprendizaje y los sistemas de evaluación, para adaptarse a las nuevas generaciones de estudiantes (FCE-UNC, 2017). En el mismo sentido, existe un compromiso institucional de la UNC y, en particular de la FCE, en que los Objetivos de Desarrollo Sostenible (ODS) constituyan un eje ordenador de sus acciones, proyectos y tareas cotidianas, reflejado en el “Programa de integración de la Agenda 2030 y los ODS en la Facultad de Ciencias Económicas” (FCE-UNC, 2022). En ese sentido, la UNESCO sostiene que “La Inteligencia Artificial (IA) proporciona el potencial necesario para abordar algunos de los desafíos mayores de la educación actual, innovar las prácticas de enseñanza y aprendizaje y acelerar el progreso para la consecución del ODS 4 (educación de calidad)” (UNESCO, s.f.). No obstante, el vertiginoso ritmo con el que avanzan las tecnologías genera nuevos riesgos y desafíos que aún no han sido suficientemente abordados por las políticas públicas ni por los marcos normativos vigentes.

En este contexto, durante el ciclo lectivo 2024, se diseñó e implementó una actividad grupal con asistencia de herramientas de Inteligencia Artificial Generativa (IAG), en la que participaron 98 estudiantes, con el objetivo de explorar el potencial de la IAG en la modelización de problemas de programación lineal en contextos complejos, promoviendo un enfoque crítico e integrado en su uso académico y profesional. Para la consecución del mismo, como objetivos específicos se pueden mencionar: i) Fortalecer la comprensión de los principios de modelización matemática y programación lineal, en articulación con un uso técnicamente fundamentado de la IAG; ii) Desarrollar una actitud crítica y reflexiva frente a las soluciones generadas por sistemas de IAG, promoviendo su evaluación sistemática y evitando su uso como una “caja negra”; y iii) Fomentar el trabajo colaborativo y el uso responsable de la IAG en contextos académicos y profesionales, promoviendo el intercambio de perspectivas y la construcción colectiva del conocimiento.

Metodología

Paradigma pedagógico

La propuesta se enmarca en un enfoque de aprendizaje activo, que sitúa a cada estudiante como agente central en la construcción de su conocimiento (Mallik & Gangopadhyay, 2023). Bonwell y Eison (1991) definen este enfoque como cualquier estrategia didáctica que involucre al grupo de estudiantes en actividades de análisis, síntesis y evaluación, más allá de la escucha pasiva. Asimismo, Prince (2004) destaca que el aprendizaje activo mejora la retención del conocimiento y favorece el desarrollo de habilidades de pensamiento crítico, lo cual lo convierte en un enfoque particularmente valioso en contextos de educación superior.

Dentro de ese paradigma, la propuesta se sustenta en los principios del Aprendizaje Basado en Problemas (ABP), una metodología centrada en el trabajo grupal y en la resolución de situaciones problemáticas que demandan comprensión profunda, reflexión crítica y generación de soluciones pertinentes (Savery & Duffy, 1995). El ABP contribuye al desarrollo de habilidades para el trabajo cooperativo, así como a la formación de competencias cognitivas de orden superior, como el análisis, la evaluación y la síntesis (Hmelo-Silver, 2004). El equipo docente, por su parte, asume un rol de facilitador, guiando el proceso más que transmitiendo respuestas (Morales & Landa, 2004).

En el marco del ABP, la Inteligencia Artificial Generativa se incorpora como un instrumento pedagógico que facilita la exploración y análisis de los problemas planteados. En ese sentido, la IAG funciona como un recurso de apoyo que fomenta la reflexión y el debate entre pares, coherente con la idea de que los estudiantes aprenden más cuando participan activamente en su proceso formativo. Esta herramienta, emergente y estratégica, posibilita la personalización del aprendizaje pero, al mismo tiempo, plantea nuevos desafíos en cuanto a la validación y el análisis crítico

de las respuestas generadas por estos sistemas (Holmes et al., 2019; Luckin et al., 2016).

La propuesta didáctica, diseñada a partir de este enfoque pedagógico, fomenta la construcción de conocimiento a través de la acción: los equipos modelan, verifican, corrigen y racionalizan sus decisiones al interactuar con IAG. La actividad permite articular conceptos teóricos y prácticos en un contexto dinámico y colaborativo, desarrollando un aprendizaje significativo mediante la experiencia directa y la reflexión crítica. En este caso, los problemas asignados no sólo ofrecen un desafío técnico, sino que promueven la discusión, el intercambio y la toma de decisiones consensuadas dentro de cada equipo. Asimismo, se busca favorecer el desarrollo de competencias que permitan al grupo de estudiantes interactuar de manera crítica con la IAG, identificando sus limitaciones y potencialidades.

Finalmente, la propuesta se alinea con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030, en particular con el ODS 4 “Garantizar una educación inclusiva, equitativa y de calidad”, al fomentar la formación de habilidades para la vida, la tecnología y el pensamiento crítico. Esta integración entre ABP, IAG y ODS facilita la formación de profesionales capaces de interactuar críticamente con tecnologías emergentes, en un contexto global comprometido con la equidad y el desarrollo sostenible.

Diseño del estudio

El estudio adopta un diseño exploratorio a partir de técnicas cualitativas, centrado en el análisis de una experiencia pedagógica situada en el contexto institucional descrito ut supra. Se trabajó con un muestreo no probabilístico de tipo accidental de $n=23$ grupos. La recolección de datos se realizó mediante un auto-reporte escrito, a partir de una encuesta grupal con preguntas abiertas, con el objetivo de relevar percepciones, valoraciones, aprendizajes y dificultades asociadas al uso de la IAG como herramienta de apoyo en la modelización de problemas de programación lineal. Las respuestas obtenidas constituyeron el corpus de análisis del estudio. En una primera etapa, siguiendo a Hernández Sampieri et al. (2014), se realizó una lectura exhaustiva de las respuestas buscando asociaciones lógicas.

Posteriormente, se efectuó una codificación, en la que se identificaron elementos y rasgos que conforman las principales reflexiones vertidas por los grupos. En esta fase, se tomó a las frases como unidades de análisis y se utilizó un método de comparación constante. Finalmente, computando los valores codificados, se definieron cuatro categorías emergentes del análisis de los datos, en diálogo con los objetivos del estudio y el marco conceptual de referencia: utilidad práctica, calidad de las respuestas, facilidad de uso y aprendizaje, y trabajo en equipo.

Limitaciones

La estrategia metodológica adoptada da cuenta de una serie de limitaciones que resulta necesario explicitar. Por el carácter exploratorio y emergente de la experiencia pedagógica analizada, no se incorporaron de manera sistemática otras dimensiones relevantes que podrían haber permitido profundizar los análisis posteriores. En particular, no se consideraron aspectos vinculados a los conocimientos previos en programación lineal, al nivel de familiaridad individual con herramientas de IAG, ni a las trayectorias académicas de los/as estudiantes participantes, dimensiones que podrían incidir en las percepciones y valoraciones expresadas. Por otro lado, se trata de una experiencia situada, desarrollada en una asignatura, carrera y contexto institucional específicos, lo que condiciona el alcance de los hallazgos. Asimismo, el tamaño de la muestra y el carácter grupal de las respuestas no permiten realizar generalizaciones estadísticas, sino aportar evidencia de tipo cualitativo orientada a la comprensión del fenómeno estudiado.

Implementación

La actividad se llevó a cabo en grupos de no más de 5 estudiantes para lo cual se utilizó el recurso “Elección de grupo” del aula virtual de la materia en plataforma Moodle, lo que les permitió decidir libremente la conformación de los equipos de trabajo.

Una vez finalizada la etapa de inscripción, las docentes asignaron a cada grupo distintos problemas de programación lineal. La consigna consistía en que los equipos copiaran y pegaran el enunciado del problema en ChatGPT y solicitaran su formulación matemática. Un aspecto central de la actividad fue la verificación y el análisis crítico de la respuesta inicial proporcionada por la IAG. Se pidió a los grupos que evaluaran si el modelo propuesto cumplía con los supuestos de la programación lineal y que justificaran cada uno de ellos. En caso de identificar incumplimientos, debían señalar los errores y reformular la consulta para corregirlos.

Además, se solicitó analizar la función objetivo y cada una de las restricciones, describiendo su significado e identificando sus unidades de medida. Luego de este proceso de internalización del problema, los equipos debían enumerar los errores detectados en la modelización inicial. A través de múltiples interacciones con ChatGPT, debían resolver dichos errores, registrando cada consulta y su correspondiente respuesta. Este proceso iterativo continuaba hasta alcanzar una formulación correcta, según el criterio del propio grupo.

Por su parte, el equipo docente ofreció instancias de acompañamiento, tanto en clases presenciales como a través de foros de intercambio en el aula virtual, a fin de orientar y supervisar el desarrollo de la actividad.

El informe final debía incluir todas las interacciones con la IAG y concluir con una reflexión sobre la herramienta y la experiencia desarrollada. Dicho informe debía ser entregado a través del “Buzón de tareas” habilitado en el aula virtual. Para la evaluación, se utilizó una rúbrica analítica que fue compartida con los grupos al inicio de la actividad, con el propósito de transparentar los criterios de corrección.

Resultados

En esta experiencia se utilizó como técnica de recolección de datos la encuesta grupal con preguntas abiertas, a partir de la cual se obtuvieron 23 respuestas grupales. Estas fueron receptadas en forma escrita y codificadas

para realizar su análisis. Dicho proceso consiste en fragmentar, conceptualizar y articular los datos en una nueva forma, dando las bases para el surgimiento de categorías de análisis bajo una perspectiva de abordaje cualitativo.

Para identificar y clasificar de manera sistemática los aspectos positivos y negativos asociados a la experiencia en general y, en particular, al uso de la IAG como herramienta de apoyo en la modelización de problemas de programación lineal, se agruparon las valoraciones en cuatro categorías: utilidad práctica, calidad de las respuestas, facilidad de uso y aprendizaje, y trabajo en equipo. En cada una de ellas, se seleccionaron citas de diferentes grupos que sintetizan los principales hallazgos de la experiencia.

Utilidad práctica

Esta categoría abarca todas las valoraciones relacionadas con la capacidad de la herramienta para asistir y facilitar la resolución de problemas en la práctica. Incluye aspectos como la rapidez para obtener resultados, la eficiencia en la formulación de modelos, la utilidad para organizar información y la posibilidad de abordar problemas con diferentes grados de complejidad.

Los grupos de estudiantes destacaron que el uso de ChatGPT agilizó y simplificó la formulación de modelos matemáticos, y demostró ser útil para organizar información, definir variables y restricciones. La rapidez en las respuestas y la utilidad para abordar problemas de baja complejidad fueron especialmente valoradas.

En ese sentido, el grupo 02 manifestó:

La IA simplifica el proceso de formulación de modelos matemáticos y restricciones, ayuda a ahorrar tiempo en la toma de decisiones. (grupo 02),

mientras que el grupo 19, compartió la siguiente reflexión:

En el desarrollo de este trabajo hemos identificado que herramientas de inteligencia artificial como el ChatGPT son de gran utilidad al agilizar el proceso de resolución y análisis de problemas. (grupo 19)

Sin embargo, la utilidad práctica se vio limitada por la necesidad de realizar múltiples ajustes y validaciones para alcanzar el modelo correcto. Así, el grupo 14 sostuvo que:

Como herramienta de apoyo, ChatGPT resulta práctica, siempre y cuando se tenga un conocimiento previo claro de lo que se desea plantear y se presente la información de manera correcta y con el máximo nivel de detalle posible. (grupo 14)

Esto muestra que, si bien es un apoyo valioso, no reemplaza por completo las metodologías tradicionales ni garantiza una modelización de manera autónoma.

Calidad de las respuestas

Esta categoría se centra en el análisis de la precisión y consistencia de las respuestas brindadas por la IAG. Incluye tanto los aciertos en la formulación de modelos como los errores en cálculos, planteo de restricciones, uso de unidades de medida, etc. También se incluyen las observaciones sobre la evolución de las respuestas a través de sucesivas interacciones y la capacidad de la herramienta para corregir errores cuando se le señala explícitamente.

Los grupos de estudiantes valoraron que la herramienta proporciona resultados organizados, claros y detallados, lo que facilita la comprensión y el seguimiento del proceso de modelización. Asimismo, destacaron la capacidad para adaptarse a correcciones y mejorar las respuestas tras sucesivas interacciones. No obstante, detectaron errores en la formulación de restricciones, la definición de variables y el uso adecuado de unidades de medida. Entre las limitaciones más relevantes, se mencionó que ChatGPT no siempre verifica los supuestos de la programación lineal, y necesita supervisión constante y validación humana para evitar errores. A pesar de estas limitaciones, se valoró que el proceso de interacción con la IAG

promueve el desarrollo del pensamiento crítico y la capacidad de cuestionar y corregir respuestas, lo que contribuye al aprendizaje activo.

En relación con esta categoría, los grupos de estudiantes manifestaron:

Algunos de los inconvenientes tienen que ver con que sus respuestas son redundantes, y cuando se le pide que reformule alguna de ellas para obtener una mejor conclusión, simplemente responde lo mismo, pero con otras palabras. No obstante, a partir de los conocimientos incorporados en el cursado pudimos ir corrigiendo los errores arrojados. (grupo 12)

Dado que ChatGPT sugiere soluciones basadas en patrones y conocimientos preexistentes, no siempre puede adaptarse perfectamente a casos muy específicos o complejos. (grupo 13)

Facilidad de uso y aprendizaje

Esta categoría reúne los aspectos vinculados a la interacción pedagógica y la accesibilidad de la herramienta como recurso de aprendizaje. Incluye valoraciones sobre la facilidad para realizar consultas, la contribución a la comprensión teórica y el estímulo al desarrollo de habilidades como la formulación de preguntas claras y la capacidad de cuestionar las respuestas, potenciando el aprendizaje activo. También contempla las dificultades que pueden presentarse cuando el usuario carece de conocimientos previos sólidos para guiar correctamente a la IAG.

Dentro de las valoraciones relacionadas con esta categoría, se presentan a continuación reflexiones de los grupos 06 y 07, respectivamente:

Utilizar ChatGPT fue una experiencia valiosa que nos enseñó a aprovechar los recursos tecnológicos actuales en la resolución de problemas basados en programación lineal. (grupo 06)

Interactuar con ChatGPT nos permitió profundizar en los conceptos de programación lineal y reforzar la comprensión de los supuestos y la aplicación de esta técnica, por lo que nos parece una buena forma de poner a prueba nuestros conocimientos. (grupo 07)

A pesar de estos beneficios, los grupos de estudiantes coincidieron en que el uso adecuado de la herramienta requiere conocimientos previos sólidos. La IAG, por sí sola, es insuficiente para garantizar la calidad de las respuestas: sin una base teórica adecuada, el usuario corre el riesgo de aceptar datos incorrectos o malinterpretar resultados. Por lo tanto, la IAG debe ser entendida como un recurso complementario al conocimiento humano y no como un reemplazo. En este sentido, el grupo 20 opinó:

En base a nuestra experiencia realizando esta actividad, notamos que el ChatGPT es una herramienta que nos brinda un respaldo en la resolución de problemas, además de brindarnos información instantánea, pero creemos que es sumamente necesario tener conocimiento del tema que estamos consultando para evitar que nos brinde información errónea. (grupo 20)

Trabajo en equipo

Esta categoría comprende las valoraciones relacionadas con el impacto de la actividad en la dinámica colaborativa. Incluye aspectos como la posibilidad de promover el trabajo en equipo, facilitar la discusión de distintas estrategias de resolución, favorecer el intercambio de ideas y estimular la construcción colectiva del conocimiento.

En cuanto al trabajo en equipo fue buena la experiencia, ya que pudimos compartir ideas y distintos puntos de vista a la hora de resolver los distintos planteos del trabajo, además de afianzar los conocimientos aprendidos hasta el momento sobre programación lineal en la materia. (grupo 14)

Entre las dificultades se destacan la distribución desigual de tareas y la dependencia excesiva de la herramienta por parte de algunos grupos.

La propuesta de trabajar las consignas de manera grupal constituye un desafío en cursos masivos en los que las prácticas educativas suelen estar centradas en actividades individuales. Desde un punto de vista pedagógico, el trabajo colaborativo promueve el desarrollo del pensamiento crítico a través del debate y la discusión de ideas diferentes al interior del grupo.

A modo de cierre, se presenta a continuación una tabla resumen de las frecuencias de las valoraciones -positivas y negativas- de los 23 grupos para cada categoría analizada (Tabla 1) como así también una nube de palabras clave (Figura 1).

Tabla 1

Resumen de las valoraciones de los grupos por categoría

Categorías	Valoraciones positivas	Valoraciones negativas
Utilidad práctica	18	5
Calidad de las respuestas	11	12
Facilidad de uso / aprendizaje	17	6
Trabajo en equipo	19	4

Nota. Elaboración propia, los valores corresponden a la frecuencia de menciones en las encuestas grupales y no deben interpretarse como datos estadísticos representativos.

Figura 1

Nube de palabras de las percepciones de los grupos



Nota. Elaboración propia.

Finalmente, cabe señalar que los errores y aprendizajes mencionados en el análisis no se corresponden con indicadores objetivos de evaluación, sino con interpretaciones construidas a partir de los discursos de los grupos, en función de las dificultades identificadas, los ajustes realizados durante la interacción con la IAG y las reflexiones finales sobre el proceso de aprendizaje.

Conclusiones

La experiencia permitió explorar el uso de la IAG como herramienta de apoyo para la modelización de problemas de programación lineal en el marco de una propuesta pedagógica basada en el aprendizaje activo y el trabajo colaborativo. La integración de ChatGPT en esta actividad no solo propició el desarrollo de competencias técnicas y metodológicas, sino que también promovió instancias de análisis crítico, reflexión y evaluación colectiva, alineándose con los ODS, y las demandas actuales de la educación superior y del mercado laboral en contextos cada vez más digitalizados.

Los resultados obtenidos a partir del análisis cualitativo de las valoraciones estudiantiles evidencian el potencial de la IAG para facilitar la formulación y comprensión de modelos matemáticos, especialmente en problemas de menor complejidad. Las dificultades identificadas —errores en las respuestas, limitaciones en el tratamiento de casos complejos, entre otros— refuerzan la necesidad de incorporar esta tecnología desde una perspectiva pedagógica que promueva la autonomía, la validación constante y el desarrollo del pensamiento crítico.

Asimismo, se evidenció que la dinámica grupal de la propuesta favoreció la colaboración y el intercambio de ideas, promoviendo la construcción colectiva del conocimiento, aspectos particularmente relevantes en los cursos numerosos, donde tienden a imponerse las actividades realizadas de manera individual.

Sin embargo, debido a la metodología aplicada y el tamaño de la muestra, los resultados deben interpretarse como un análisis contextualizado que contribuye a la reflexión sobre prácticas pedagógicas innovadoras en educación superior, más que como conclusiones extrapolables de manera directa a otros contextos. No obstante, la experiencia permite identificar orientaciones generales para el diseño de propuestas similares en cursos de programación lineal u otras áreas afines, especialmente la necesidad de integrar el uso de IAG en un diseño didáctico explícito, con acompañamiento docente y espacios de reflexión crítica.

En síntesis, esta experiencia permite concluir que la incorporación estratégica de herramientas de IAG en propuestas didácticas cuidadosamente diseñadas puede enriquecer los procesos de enseñanza-aprendizaje, siempre que se promueva su uso responsable y ético, orientado a la formación de profesionales críticos, reflexivos y capaces de enfrentar los desafíos de un entorno académico y laboral atravesado por la transformación tecnológica.

Referencias

- Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom* (ASHE-ERIC Higher Education Report No. 1). The George Washington University, School of Education and Human Development.
- Facultad de Ciencias Económicas, Universidad Nacional de Córdoba. (2017). *Plan de Desarrollo Institucional* 2017-2021. <https://www.eco.unc.edu.ar/files/comunicacion/PDFs/LibroPDI-baja.pdf>
- Facultad de Ciencias Económicas, Universidad Nacional de Córdoba. (2022). *Resolución (RD) 1076/2022: Programa de integración de la Agenda 2030 y los ODS en la FCE-UNC*. https://www.eco.unc.edu.ar/files/comunicacion/PDFs/2022/RD-1076_2022.pdf
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6a ed.). McGraw-Hill.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235–266. <https://doi.org/10.1023/B:EDPR.0000034022.16470.f3>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Mallik, S., & Gangopadhyay, A. (2023). Proactive and reactive engagement of artificial intelligence methods for education: A review. *Frontiers in Artificial Intelligence*, 6, 1151391. <https://doi.org/10.3389/frai.2023.1151391>
- Morales Bueno, P., & Landa Fitzgerald, V. (2004). Aprendizaje basado en problemas. *Theoria*, 13, 145-157.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223–231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>
- Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, 35(5), 31–38.
- UNESCO. (s.f.). *La inteligencia artificial en la educación*. <https://www.unesco.org/es/digital-education/artificial-intelligence>

Estatuto ético y antropológico de las inteligencias artificiales

Ethical and anthropological status of artificial intelligence

Eduardo Gutiérrez Gutiérrez

egutierrezg@uemc.es

Universidad Europea Miguel de Cervantes

ORCID ID 0000-0001-7809-6530

Resumen: Las inteligencias artificiales forman parte de la vida cotidiana de cada vez más ciudadanos de todo el mundo: desde el estudiante que las utiliza para hacer los deberes y la anciana que ataja con ellas su soledad, hasta el empresario que implanta robots para mejorar la productividad de su fábrica o el oficinista que utiliza un chatbot para las gestiones diarias.

Que forman parte de “nuestro mundo” es un hecho indiscutible. La cuestión, por recordando el famoso libro de Scheler, es la de su puesto en ese mundo, es decir: ¿Qué son las inteligencias artificiales en el marco de “lo humano”?

En este artículo diseñamos una “teoría de teorías” para analizar la cuestión sobre el estatuto ético y antropológico de las inteligencias artificiales. Nos parece que esta es la primera piedra en el camino hacia una regulación normativa de las relaciones entre las inteligencias artificiales y las personas humanas.

Palabras clave: Cosas; dignidad; ética aplicada; inteligencia artificial; persona.

Abstract: Artificial intelligence is becoming part of everyday life for more people around the world: from students who use it to do their homework and elderly people who use it to combat loneliness, to business owners who implement robots to improve factory productivity and office workers who use chatbots for daily tasks.

That it is part of “our world” is an indisputable fact. The question, recalling Scheler's famous book, is its place in that world, that is: What is artificial intelligence in the context of “the human”?

In this article, we design a “theory of theories” to analyze the question of the ethical and anthropological status of artificial intelligence. We believe this is the first step on the road to regulatory norms governing the relationship between artificial intelligence and human beings.

Keywords: Things; dignity; applied ethics; artificial intelligence; person.

Introducción

Podemos afirmar que durante el primer cuarto del nuevo milenio las inteligencias artificiales se han convertido en un nuevo tema de debate dentro de la filosofía moral. Es este un evento inédito, pero análogo a otros acontecidos durante el siglo XX: irrupción del animalismo, ecologismo.

En efecto, el potente desarrollo que en las últimas décadas han experimentado las inteligencias artificiales suscita diversos problemas para la filosofía moral. Al efecto de clasificarlos de una manera probablemente grosera, pero funcional a los propósitos de esta introducción, utilizaremos la distinción que el profesor Vincent C. Müller establece para la entrada sobre ética de la inteligencia artificial de la famosa enciclopedia de filosofía Stanford. Distingue Müller entre “preocupaciones éticas” y “desafíos éticos”. *Preocupaciones* éticas suscitadas por la irrupción de las inteligencias artificiales son, por ejemplo, el problema para nada nuevo de la posible pérdida o

destrucción de puestos laborales, el de la seguridad digital o la manipulación de noticias e informaciones.

Las discusiones *mundanas* sobre los problemas implicados por las inteligencias artificiales caen a nuestro juicio en el grupo de las “preocupaciones ética”“, como también muchas discusiones académicas cristalizadas en libros y *papers*. Por ejemplo, el comentario a la ética de las IA que Russell y Norvig hacen en el capítulo 28 de *Artificial Intelligence: A Modern Approach* sobre las implicaciones filosóficas de estas tecnologías se reduce a la exposición, al modo de una lista de lavandería, de las preocupaciones éticas que suscitan las IA. Lo mismo sucede en “Ética y filosofía de la inteligencia artificial: debates actuales” de Santana-Soriano (2023) o “Artificial Intelligence as an Essential Tool for the Development of Medicine Requirement” (Khan et al., 2022).

Por poner un contrapunto y restarle de paso originalidad a este artículo, Mark Coeckelbergh reflexiona sobre el estatuto ético y antropológico de las IA, que denomina “político”, en el capítulo 6 de *La filosofía política de la inteligencia artificial* (2023). También se abordan estas cuestiones en el libro de Carlos Madrid Casado *Filosofía de la Inteligencia Artificial* (2024), escrito desde las mismas coordenadas filosóficas que este modesto trabajo.

Si bien no pensamos que sean irrelevantes, en este artículo queremos centrar la atención en el que probablemente sea el gran “desafío ético” que las inteligencias artificiales suscitan para la filosofía moral de nuestro tiempo: el problema de su estatuto ético y antropológico. Por explicarlo en cuatro palabras: ¿las inteligencias artificiales son *personas*? Es decir, ¿son *sujetos morales*? Si lo son, ¿en calidad de *pacientes* o de *agentes* morales?

Este desafío sirve para distinguir el problema moral de las IA de otros problemas tecnológicos, como el de la dualidad axiológica en el uso de la energía nuclear, capaz de lo mejor y lo peor para la especie humana. Hay, está claro, implicaciones éticas en el uso *técnico* que puede hacerse de la energía nuclear, pero son, por decirlo de alguna manera, de menor calado que las implicaciones éticas del uso de las inteligencias artificiales. Nadie en su sano

juicio se atreverá a discutir la condición moral o personal de la energía nuclear, pero es lícito hacerlo respecto de las IA.

La cuestión sobre el estatuto ético y antropológico de las inteligencias artificiales no sólo es central para una ética de la IA que quiera estar a la altura de los tiempos. Es, además, un verdadero desafío para la filosofía moral en general, porque nos pone ante el problema eludido por muchos de si no será necesario revisar nuestro sistema de conceptos éticos, esto es, la tradición ética tal y como la conocemos, de modo análogo a como lo propuso en 1993 el Proyecto Gran Simio de Singer y Cavalieri. Léase, por ejemplo, el artículo “En defensa de la condición de persona de los gorilas” de Francine Patterson y Wendy Gordon (1998).

A nuestro juicio, la literatura sobre la ética de la IA parece eludir las implicaciones antropológicas del problema del estatus moral de estas nuevas tecnologías. Y nos parece que, a tenor de los posibles cursos de desarrollo de las IA, que según los expertos apuntan a la aparición de una “superinteligencia” con aspecto indistinguible del humano, las implicaciones antropológicas son las más acuciantes. No se trata solamente de determinar si pueden actuar autónomamente o no. De lo que se trata, lo que tiene verdaderas repercusiones en la filosofía moral, es si debemos incluirlas o no dentro del “círculo de reconocimiento moral”.

Como siempre, a la quizás un tanto grosera afirmación que abre el párrafo anterior se pueden alegar contraejemplos como el del número especial 12 que la revista *Naturaleza y Libertad. Revista de Estudios Interdisciplinarios* dedica en el año 2019 a la relación entre la antropología filosófica y la Inteligencia Artificial, bajo el rótulo *¿Es posible transferir la mente humana a un soporte no biológico?* O también, el importante artículo “¿Es posible transferir la mente humana a un soporte no biológico?” (Linares-Peralta & Llorca Albareda, 2024), donde, en términos semejantes a los nuestros, plantean la necesidad de repensar las categorías tradicionales de la antropología filosófica.

No resulta difícil deslizarnos de la cuestión sobre el estatuto ético de las inteligencias artificiales a la de su estatuto antropológico, de manera semejante a como se hizo con los animales no humanos: a tenor de las

características humanoides que comienzan a manifestar, como la inteligencia o el aprendizaje, ¿no cabría considerar a las inteligencias humanas como personas? Si no lo son, ¿cómo justificar la aproximación ética a estas tecnologías? ¿No sería, más que un problema ético sensu stricto, un problema técnico, a saber, el de la repercusión que el uso con las tecnologías de inteligencia artificial tiene sobre las vidas de otros sujetos morales? Y si lo son, ¿en qué se diferencian de los seres humanos?

En este artículo nos proponemos un análisis filosófico del estatuto ético y antropológico de las inteligencias artificiales como prolegómeno a una ética de la inteligencia artificial. Como hipótesis de trabajo que ahora no es pertinente justificar consideramos necesario, para orientarnos por estos derroteros, partir de un determinado sistema de filosofía. En efecto, como estamos viendo en esta introducción, para responder a la pregunta central de la ética de la inteligencia artificial precisamos de una filosofía política y de una antropología filosófica, por no hablar de una filosofía de la ciencia, de una ontología y hasta de una filosofía de la mente (¿son *inteligentes* las inteligencias artificiales?, etc.). Nuestra toma de partido es por el sistema del materialismo filosófico de Gustavo Bueno, del que extraeremos el instrumental necesario para la prospección deseada.

A continuación, realizamos una breve introducción a la ética de la inteligencia artificial. Después, planteamos sucintamente los parámetros de la filosofía moral del materialismo filosófico, especialmente el artificio conceptual del “espacio antropológico”¹. En la segunda parte del trabajo localizamos el lugar que las inteligencias artificiales ocupan en el “espacio antropológico”, concluyendo con algunos comentarios sobre su estatuto ético y una síntesis de la propuesta conceptual.

¿De qué hablamos cuando hablamos de “ética de la inteligencia artificial”?

Desde la sistematización de eso que llamamos filosofía moral en la Atenas de los filósofos, el ámbito de la moralidad, que llamaremos “círculo de reconocimiento moral”, solamente protegía, dotando de derechos y *dignidad*, esto es, reconociendo como sujetos morales, al conjunto de los sujetos

operatorios humanos. Ese CRM actúa, por utilizar una metáfora luminosa, como un *paraguas* que protege del temporal de violencia y abusos que los hombres practican sobre otros hombres: a quien está bajo ese paraguas no se le puede dañar porque, por decirlo con Kant, es un fin en sí mismo y no solamente un medio; posee *dignidad* y no solo *precio*.

Es cierto que hasta hace poco tiempo el CRM solamente cubría a los seres humanos, pero no lo es que los cubría a todos sin excepción (ni siquiera hoy lo es: inmigrantes y refugiados se encuentran desprotegidos de los derechos humanos a consecuencia de la ausencia de derechos civiles y políticos). En Grecia y Roma amplias capas de la población eran expulsadas de ese círculo, como los esclavos, los extranjeros, las mujeres, los niños o los discapacitados mentales.

La declaración universal de los derechos del hombre y el ciudadano de los jacobinos revolucionarios supondría la primera propuesta positiva de ampliación del círculo a la totalidad de los seres humanos (con los antecedentes del estoicismo y de la escolástica española). A partir de este momento, el círculo de reconocimiento moral incluirá a (todas) las personas humanas bien en calidad de agentes morales (a quienes se les reconoce autonomía *moral*, es decir, responsabilidad en la toma de decisiones que afectan a otros sujetos humanos, y por consiguiente deberes u obligaciones para con los otros), bien en calidad de pacientes morales (a quienes no se les puede reconocer deberes, pero sí derechos y dignidad).

En el siglo XX, las diversas corrientes de ecoética (o *ecoéticas*) fueron ampliando progresivamente el radio del círculo de reconocimiento moral más allá de los límites antropológicos impuestos por la tradición filosófica (antropocentrismo en sus versiones fuerte y débil): sensocentrismo, biocentrismo, ecocentrismo, ética de la Tierra, etc. Durante el primer cuarto del siglo XXI el debate concerniente a lo que desde hace casi un siglo se llama “ética aplicada” gira en torno a si se deben introducir o no en ese círculo de reconocimiento moral nuevas entidades no humanas ni animales que sin embargo parecen dar muestras de la posesión de facultades y capacidades humanas: las inteligencias artificiales.

El de la ampliación del círculo de reconocimiento moral es a nuestro juicio el problema central de la ética de la inteligencia artificial (sin obviar, desde luego, la importancia de otros que en la introducción hemos llamado “preocupaciones éticas”), referente a su estatuto ético y antropológico: ¿cómo debemos considerar o concebir a las tecnologías de inteligencia artificial? ¿Nuestro trato con las inteligencias artificiales debe ser un trato ético? ¿Las inteligencias artificiales tienen derechos? ¿Y obligaciones? Este problema, aun moral, como decimos, también reverbera en la esfera antropológico-filosófica como el problema de la discontinuidad o continuidad entre seres humanos y máquinas (Linares-Peralta & Llorca Albareda, 2024, p. 4).

Problemas de la ética de la inteligencia artificial

La ética de la inteligencia artificial, como parte de una filosofía de la inteligencia artificial que por necesidad debe incluir una filosofía de las ciencias y una ontología, se ocupa de problemas muy diversos que podemos clasificar en tres grandes grupos:

Ética en el diseño de las IA

En los últimos años las tecnologías de inteligencia artificial se están aplicando a cada vez más campos: trabajo (empresa china de Shenzhen), recursos humanos (valoración de currículos para la ocupación de vacantes laborales), cuidados sociales, vigilancia y control policial, defensa y armamento militar, resolución de dudas, generación de textos e imágenes, producción de noticias, etc. Implicados a estas sucesivas aplicaciones encontramos problemas éticos como el de la manipulación o la desinformación (lindantes con la filosofía del periodismo, en el sentido del genitivo subjetivo), de los sesgos ideológicos y la discriminación, la ciberseguridad y el robo de información, entre otros. Más que problemas, hablamos de preocupaciones, de acuerdo con lo mencionado al inicio.

El problema que sirve como hilo rojo para conectar a los que hemos expuesto en el párrafo anterior es el de la *responsabilidad*: ¿a quién debemos exigir responsabilidades o rendición de cuentas? ¿A las propias tecnologías de IA? Esto nos lleva al tercer grupo de problemas que vamos a clasificar. ¿A los

inversores de las compañías productoras de esas tecnologías? ¿A los usuarios?, ¿a sus entrenadores o diseñadores?

Este es el principal foco de problemas que se está discutiendo hoy en día, fundamentalmente desde el Derecho, estableciendo límites legales en el uso y diseño de las IA en base a principios de transparencia, seguridad, justicia y no discriminación, respeto de la autonomía humana, responsabilidad y rendición de cuentas. Digamos, por consiguiente, que este primer grupo de problemas se encuentra a medio camino entre la ética de las inteligencias artificiales y el Derecho: (Cerrillo, Martínez & Peguera Poch, 2020), (Garapon & Lassègue, 2021), (Herrera Triguero, Peralta Gutiérrez & Torres López, 2022).

Ética de las IA (o para las IA)

Toda vez que reconocemos el grado de autonomía que las IA de última tecnología están adquiriendo para la toma de decisiones con poderosas implicaciones antropológicas, ¿no sería conveniente introducirles algorítmicamente sistemas éticos que regulen esa toma de decisiones?

Según explican Anderson y Anderson (2007), la ética de las máquinas refiere a un ámbito de análisis filosófico moral que reflexiona sobre el comportamiento que las IA, o las máquinas en general, tienen hacia sus usuarios humanos. Por ejemplo, el comportamiento que un robot programado para cuidar de un anciano en una residencia tiene hacia este.

El ejemplo que suele utilizarse como hipótesis para este tipo de discusiones es el del coche autónomo que se enfrenta a una de esas situaciones que llamamos “dilemas éticos”, en la que se plantean diversos cursos de acción entre los que es preciso elegir justificada o razonablemente: ¿cuál es la decisión que *debería* tomar la máquina? En la novela de Asimov *I, robot* esto se ve claramente: si una máquina debe decidir entre salvar la vida de una niña pequeña y la de un hombre adulto de mediana edad, ¿a quién elegiría salvar y por qué? Y lo que es más importante: ¿Podría justificar su toma de decisión aportando argumentos moralmente significativos (válidos para todos)?

El problema anterior no parece revestir dificultades reseñables. Mayoritariamente se convendrá en la necesidad de introducir un sistema ético que facilite la toma de decisiones de las máquinas en favor de los humanos. Sin embargo, ¿cuál de los muchos sistemas éticos realmente existentes en la tradición filosófica, y entre los cuales hay incompatibilidades dialécticas, es el apropiado para este diseño normativo?, ¿quién tiene la autoridad para decidir? ¿Es mejor el deontologismo o el utilitarismo? ¿Ayudan estos sistemas de la tradición a tomar decisiones en un mundo como el nuestro? A juicio de Brožek y Janik (2019), es más fácil para una máquina que para un ser humano atenerse a las máximas de sistemas éticos tan abstractos como el deontologismo kantiano o el utilitarismo de Bentham. Esta paradoja dice poco en favor de esos sistemas morales, cuya idealización los aleja completamente de la práctica moral realmente existente.

El primer sistema de normas éticas (no ya jurídicas) para el diseño de las Inteligencias Artificiales lo constituyen las tres famosas leyes de la robótica de Asimov (*I, Robot*, 1950): Un robot nunca hará daño a un ser humano ni permitirá que un ser humano sufra daños; Un robot deberá obedecer las órdenes que le dé un ser humano, salvo en los casos en que tal obediencia contravenga la primera ley; Un robot deberá proteger su propia existencia, siempre que dicha protección no entre en conflicto con las otras dos leyes.

Ahora bien, es discutible que un sistema de normas que se implantan en un sujeto operatorio inanimado durante su proceso de diseño matemático e informático para que actúe en conformidad a ellas, como si se tratara de pautas de conducta más que de normas en el sentido estricto del concepto, pueda ser llamado un sistema ético.

Para que podamos hablar de sistema ético, o de ética en sentido amplio, se requiere de la autonomía del agente moral como condición necesaria. A grandes rasgos, distinguimos dos tipos de autonomía: autonomía cognitiva (conocimientos para realizar alguna operación compleja de manera independiente, sin asistencia) y autonomía moral (responsabilidad, conciencia de los motivos y las consecuencias de nuestras acciones). Un ejemplo para captar la diferencia entre ambos tipos: un individuo al que se le expide el carné

de conducir ha demostrado poseer autonomía en el sentido cognitivo para la práctica de conducción de un automóvil. Sin embargo, ese mismo individuo, por descuido, puede atropellar a una persona que cruza un paso de cebra y demostrar falta de autonomía moral si decide no asistirle y huir, no haciéndose con ello cargo de las consecuencias de su acción.

Si no hay autonomía (en el sentido moral), no hay ética más que como cuidado hacia el paciente moral. En todo caso, no tenemos agente moral que actúa conforme a normas. Así las cosas, parece que las inteligencias artificiales no son *autónomas* sino *automáticas*, y las normas en virtud de las cuales se las entrena no son normas éticas sino comandos de acción semejantes a los que el usuario de videojuegos transmite a su avatar digital.

En este grupo de problemas cabe incluir la reflexión sobre la incidencia que una extensión en el uso con las IA puede tener sobre las relaciones humanas, en términos de pérdida de habilidades cognitivas (Greenemeier, 2011) y personales (Groth, Nitzberg & Esposito, 2018), o de erosión de la sociedad civil (como se demostró en el año 2018 con el escándalo Cambridge Analytica-Facebook).

Ética con las IA: la cuestión de la “agencia moral”

En nuestro día a día *tratamos*, ya no simplemente *utilizamos*, con las inteligencias artificiales. La batería de problemas clasificados en los grupos anteriores advierte sobre los peligros que este trato puede entrañar para los seres humanos, mientras este tercer grupo nos pone ante el espejo de nuestras propias acciones y conductas *morales*: ¿debemos considerar a las IA como sujetos morales? Es decir, ¿debemos regular normativamente nuestro trato con las IA, como un trato moral con entidades personales?

Según informan algunos diarios haciéndose eco de recientes y divertidos memes, hay un porcentaje considerable de usuarios que tratan cortés o educadamente a las IA con las que interactúan diariamente (chatbot generalmente) por miedo a que se vuelvan hostiles y peligrosas en el futuro.

Esto remite al argumento del antropologismo débil defendido por filósofos como Santo Tomás o Kant: hay que regular el trato entre los seres humanos y los seres no humanos, no por el valor intrínseco que les reconocemos a los animales sino por su valor *para sí*, para la especie humana: el ejercicio de la violencia contra un animal podría llevarnos por la pendiente resbaladiza de la violencia hacia otros humanos. *Mutatis mutandis*, en este caso tratamos bien o moralmente (educadamente) a las máquinas no porque las consideremos sujetos morales, sino por miedo a que el día de mañana se rebelen contra nosotros.

La urgencia de este problema puede ponerse en comparación con el despertar de la conciencia animalista. Hasta el siglo XIX se pensaba que entre los seres humanos y los animales existía un abismo teológico, perfectamente fundamentado por el mecanicismo de raigambre cartesiana: los animales son pura *res extensa*, máquinas que se mueven instintivamente a la manera de autómatas, mientras los seres humanos son *res pensante*, entidades con espíritu. La causa de ese *gorismós* es la insuflación del alma en el cuerpo animal de los hombres por la vía divina.

La teoría darwiniana de la evolución fue el punto de inflexión que comenzó a tambalear el dogma mecanicista, aduciendo una continuidad biológica entre animales y humanos. La genética moderna vendría a confirmar la ruptura del hiato y a demostrar una semejanza inaudita entre los humanos y nuestros parientes zoológicos.

Huelga decir que la filiación biológica, genética o fenotípica entre los humanos y las tecnológicas de IA es imposible. Sin embargo, hay una analogía entre ambos casos que se puede explicar por vía de la intuición empírica. Antes de Darwin y de las teorías predecesoras del darwinismo, se intuía una semejanza conductual entre los animales y los humanos. Así, por ejemplo, el sensocentrismo: si los animales son capaces de sentir, ¿por qué no incluirles como miembros de pleno derecho en el círculo de reconocimiento moral? Dicha intuición empírica se puede aplicar perfectamente al caso de las inteligencias artificiales, que no en vano son consideradas efectivamente como inteligentes (otro tema que no nos ocupa aquí es el de si de hecho lo

son, y qué es la inteligencia). Si las IA razonan, aprenden y piensan, como lo hacemos los humanos, ¿por qué no incluirlas también en el círculo de reconocimiento moral?

En caso de aceptar esa inclusión, y de reconocer que las IA son sujetos morales, ¿en calidad de pacientes o de agentes morales? Son autónomos y toman decisiones, luego parece que son más agentes que pacientes morales.

Para salvar la pendiente resbaladiza por la que amenaza conducirnos esa última declaración se ha acuñado la categoría de “agentes morales funcionales”, referente a sujetos morales que pueden tomar decisiones significativas, es decir, con consecuencias que afectan a la vida de las personas humanas, pero sin intencionalidad. Van Wynsberghe y Robbins (2019) utilizan el concepto de “agentes morales autónomos”. Otros, como Allen, Varner y Zinser (2000), prefieren el de “agentes morales artificiales” para limitar las expectativas de su compromiso.

Mandy Zafar (2024) niega que las IA sean agentes morales, y propone el concepto de “AI automated performance” (AIAP). Viene a decir que en las discusiones recientes se utiliza el concepto de agente moral de una manera demasiado laxa y simplista, atribuyéndoles a las tecnologías de IA tal estatus moral en virtud de su autonomía, racionalidad e inteligencia. El concepto de agente moral se utiliza preferentemente para los seres humanos, aunque por analogía lo empleemos también para interpretar la conducta de otros seres no humanos como nuestras mascotas. Dice a este respecto:

Llamar “agentes” a los animales o decir que son capaces de tener acciones voluntarias no significa automáticamente que sean agentes morales. Ser considerado un “agente moral” conlleva repercusiones normativas adicionales, entre ellas la atribución de capacidades jurídicas o la recepción de derechos legales. Sin embargo, en algunos casos se atribuyen derechos legales sin una clara atribución de agencia moral, como ocurre, por ejemplo, con los animales. Existe un amplio consenso en que los animales son al menos receptores de bienestar, pero no actores de voluntad propia. Esta es una cuestión abierta al

debate. Aunque estos temas son muy interesantes e importantes, no se pueden tratar con más detalle aquí. (Zafar, 2024, p. 2)

James Moor (2006) clasifica cuatro tipos de agentes morales: agentes de impacto ético (un conductor robot), agente ético implícito (un piloto automático), agente ético explícito (capaz de realizar juicios en base a la utilidad de cada alternativa) y agente ético completo (que hace juicios éticos y es capaz de justificarlos). La pregunta es: ¿Se pueden crear agentes éticos completos, categoría hasta ahora solamente reservada para los seres humanos, *artificialmente*?

Además, puestos a utilizar la idea de autonomía como criterio para la concesión de agencia moral, convendría distinguir dos concepciones: una cosa es la autonomía como saber hacer algo, o autonomía cognitiva, y otra la autonomía en el sentido moral, ligada a la responsabilidad. Cuando a alguien se le expide el carné de conducir es porque ha demostrado ser capaz de conducir autónomamente sin necesidad de un profesor acompañante en el asiento del copiloto. Decimos que ese conductor es autónomo o tiene la capacidad de conducir autónomamente. Sin embargo, si ese conductor una noche va despistado y por saltarse un paso de cebra atropella a una persona y huye, incurriendo en un delito de omisión de socorro, demuestra no tener autonomía moral, esto es, responsabilidad, capacidad para hacerse cargo de las consecuencias de sus acciones.

El problema de la autonomía de las inteligencias artificiales, tal y como lo estamos exponiendo, está directamente relacionado con el de su racionalidad.

Pero a nuestro juicio el de la racionalidad no es un criterio suficiente para la inclusión de un sujeto en el círculo de reconocimiento moral. La etología ha demostrado que todas las especies animales son racionales (“raciomorfas”, se dijo durante una época para marcar las diferencias), pero no es la racionalidad lo que las incluye en dicho círculo; de hecho, muchas de ellas ni siquiera son incorporadas, como los insectos. Ninguna de las ecoéticas que reivindican la incursión de las especies animales en el círculo de reconocimiento moral apelan a la racionalidad como criterio justificativo. Es

más, la racionalidad fue el criterio que en tiempos sirvió para dejar fuera del círculo de reconocimiento moral a discapacitados mentales, niños, ancianos dementes y extranjeros (sin *logos*).

En su conocidísimo manual *Artificial Intelligence: a modern Approach* Russell y Norvig ofrecen una definición de la acción racional que a nuestro juicio resulta insatisfactoria: “Un agente racional es aquel que actúa de manera que logre el mejor resultado o, cuando hay incertidumbre, el mejor resultado esperado” (2026, p. 22).

Esa distinción, que ejerce una idea de racionalidad teleológica, no discrimina los *finis operantis* a los que se propone llegar el agente racional de los *finis operis* a los que la acción llega objetivamente, con independencia de que el agente racional buscase o no tal fin. Lo que nos interesa en esta ocasión es la preparación proléptica del *finis operantis*, porque en la ejecución de la acción pueden influir diversos factores capaces de alterar su curso. La estrategia racional es, en efecto, proléptica, es decir, proyectada hacia el futuro, pero en toda prolepsis se implica una anamnesis o recuerdo en el que se arrastran normas institucionalizadas (y socializadas) que contribuyen a la preparación de la estrategia. Por lo tanto, la racionalidad, más que en las ideas internas del sujeto racional, está en las instituciones y las normas segregadas de estas en las que ese se apoya para el cumplimiento de *sus finis operantis*. A juicio de Bostrom y Yudkowsky (2014), una inteligencia artificial no tiene propósitos o fines en el sentido de *finis operantis*. Han sido sus programadores quienes las crean en aras a dichos objetivos, en el sentido de *finis operis*.

Como los propios autores comentan en el capítulo 28 del libro, el diseño computacionalista de las IA, lo que se llama Good Old-Fashioned AI, es absolutamente incapaz de replicar esa dimensión normativo-institucional fundamental para explicar la racionalidad humana. Es lo mismo que defiende el “argumento de la informalidad del comportamiento” de Turing: “el comportamiento humano es demasiado complejo como para ser capturado por un conjunto formal de reglas [lógicas, binarias, simples]” (Russell y Norvig, 2023, p. 1033). Los sistemas de razonamiento probabilístico y el *deep*

learning podrían, a su juicio, lograr la replicación más o menos fiel de una racionalidad compleja no sujeta a normas lógicas o silogísticas.

En el siguiente apartado esclarecemos cuál es nuestra definición de la “agencia” o el “agente” moral.

Por último, también cabe analizar los problemas éticos derivados del papel que las IA pudieran desempeñar en la vida buena de los seres humanos, que según Sócrates es la pregunta fundamental de la ética. Sin embargo, el de la “vida buena” o “lo mejor” es un criterio muy poco apto para las reflexiones concernientes a la ética de la inteligencia artificial porque no queda del todo claro qué vida es la buena ni para quién lo es. Además, las virtualidades para la vida buena que las IA ofrecen por ejemplo a los clientes de una sucursal bancaria suiza que ha implantado esta tecnología para agilizar y dinamizar su servicio de atención al cliente pueden ser contraproducentes o irrelevantes para un solicitante de hipoteca de raza negra (según el ejemplo con el que abren el capítulo 14 de su libro Bostrom & Yudkowsky, 2014).

Filosofía moral del materialismo filosófico

En este apartado hacemos una breve reexposición de la filosofía moral del materialismo filosófico, basándonos en el libro de 1996 *El sentido de la vida*.

Moralidad y normas

Definimos la “moralidad”, en un sentido amplio, como el conjunto de las normas (éticas, morales y jurídicas) que regulan las relaciones interpersonales; es decir, las relaciones mantenidas entre sujetos operatorios humanos a quienes se les reconoce la condición ética y jurídica de personas.

La idea de persona tiene como contexto genético-operatorio la escena teatral griega, concretamente, la máscara que los actores portaban para salir a escena y representar a su personaje (prosopón, per sonare). En base a este origen tecnológico, concebimos la categoría de persona más como una institución sociohistórica que como un atributo inherente a los seres humanos:

Persona humana añade algo no sólo a “persona” sino también a “humano”. *El hombre recibe una determinación importante cuando se le considera como persona así como la persona recibe una determinación no menos importante cuando se la considera como humana.* Por tanto, no es lo mismo hombre que persona, como tampoco es lo mismo hombre que ciudadano. (Bueno, 1996, p. 115)

La persona es una institución con una historia que se abre paso en sociedades desarrolladas con un lenguaje articulado de pronombres personales y con leyes que protegen la propiedad privada (de familias, más que de individuos). Sólo dadas estas condiciones es posible la operación jurídica de reconocimiento de la condición de persona a un sujeto operatorio, siempre en el marco de una comunidad de personas. Y tal como una persona recibe ese atributo histórico por vía institucional, puede perfectamente serle retirado (considerándole “persona cero”). Como no podemos extendernos más en este punto, remitimos al artículo de Gustavo Bueno “Para una construcción de la Idea de Persona” (1953).

La idea central de la filosofía moral del materialismo filosófico es la de Norma (en un sentido antropológico, más que jurídico). Gnoseológicamente, las normas desempeñan en el ámbito de las ciencias humanas un papel análogo al de las leyes naturales en las ciencias naturales. La diferencia es que las leyes normativas pueden ser justificadas y desobedecidas, mientras las leyes naturales pueden ser explicadas y no desobedecidas, pero sí controladas.

Los códigos normativos son conjuntos de instrucciones que han cristalizado al calor de ciertas instituciones antropológicas y regulan la ejecución de procesos operatorios realizados sobre entidades humanas o no humanas prolépticamente, esto es, con vistas a la consecución de determinados fines referidos a la supervivencia y progreso de un grupo social. La consolidación de una norma para la praxis humana se produce cuando en la confluencia de diferentes rutinas operatorias habilitadas para configurar objetos o situaciones repetibles, como la caza o el fuego, unas prevalecen sobre otras que resultan vencidas aunque se mantienen como virtualmente

posibles. Las normas, por decirlo en cuatro palabras, son rutinas victoriosas. Además, nos sirven como criterio para distinguir, rompiendo con la distinción clásica, la *ética* de la *moral*.

Ética y moral

El materialismo filosófico se opone a la interpretación tradicional (presente en todos los manuales de ética) entre la ética y la moral, según la cual la moral es un sistema normativo que rige para una comunidad y la ética la reflexión de segundo grado sobre esa praxis moral (filosofía moral o *meta-moral*). La famosa distinción escolástica entre la *ethica utens* y la *ethica docens*.

Dicha distinción está fundada en una oposición falsa, que maneja ideas oscuras y confusas, entre teoría y práctica. En la práctica de un mecánico hay tanta teoría como en los apuntes de un estudiante de ingeniería, y en las teorías científicas están implicadas las prácticas operatorias (con *aparatos*) de los experimentos. En resolución, en la praxis moral hay teoría o conocimientos sobre las acciones morales, lo mismo que en la teoría moral, por especulativa que sea, hay implicaciones prácticas.

Además, Bueno intuye un interés gremial por justificar la necesidad de un cuerpo de expertos en ética cuya función es la de enseñar a los alumnos a pensar éticamente, como si no supieran hacerlo ya de facto. La filosofía moral académica, de los filósofos y los profesores de filosofía, procede *in medias res* de la filosofía moral mundana o el conocimiento teórico y práctico que toda persona posee acerca de la moralidad en el sentido amplio.

En efecto, sobre la “moralidad” en el sentido amplio antes comentado hay sin duda un saber mundano, una “moralidad mundana” o “sabiduría moral mundana” que incluye una práctica moral y un conocimiento sobre esa práctica moral. Como afirmó Sócrates por pluma de Platón, todos los ciudadanos poseemos ciertas nociones sobre la praxis moral en virtud de las cuales tenemos la capacidad para emitir juicios en los que valoramos las conductas y acciones propias y ajenas. Si un individuo observa un ataque violento contra un paciente moral (niño, mujer, anciano, hombre, perro... ¿robot?) sentirá rabia, enfado o hasta tristeza, y juzgará que tal acción

interpersonal es reprobable. “No se debe actuar de este modo”, dirá, queriendo con Kant que su máxima se eleve a la condición de ley universal. Probablemente mediará, aun poniendo en riesgo su propia salud física, para detener el ataque. Y eso, pensarán otros, es una acción virtuosa.

Así, por traer esta cuestión a nuestro objeto de análisis, cuando un espectador anónimo (que no es profesor de filosofía ni ha recibido formación filosófica más allá de la formación básica de la enseñanza media) ve en vivo y en directo o a través de las pantallas a un individuo o a un grupo de individuos golpear deliberadamente (quizás para evaluar su respuesta, en una suerte de experimento científico) a un robot, experimentará esas emociones que son indicio psicológico de una ofensa moral y podrá en todo caso emitir un juicio sobre el valor moral de dicha acción.

Si preguntamos a un grupo de estudiantes de bachillerato, de Universidad, o a una selección al azar de profesionales de diversos ámbitos, tendrán, sin duda, una opinión bien formada acerca del estatus moral y antropológico de las IA. Estas consideraciones, por prosaicas que sean, deben ser recogidas y analizadas críticamente por el filósofo profesional o académico.

Este saber moral mundano es semejante al saber práctico del que hablaba Aristóteles, la prudencia o *prhonesis*, distinto del saber especulativo o ciencia (*episteme*). Incluye la *sindéresis*, es decir, el conocimiento de los primeros principios prácticos. Aunque como apunta Bueno contra Kant el saber moral mundano no es universal sino plural, referido a la diversidad de pueblos realmente existentes, cabe la posibilidad de reconocer un puñado de principios morales que si no son de hecho universales podrían considerarse de tal naturaleza al menos virtualmente.

Por ejemplo, la regla de oro (no hagas para los demás lo que no quieras para ti), o, por lo menos en la tradición de la ética occidental contemporánea, los cuatro principios de la bioética (autonomía, beneficencia, no maleficencia, justicia). Efectivamente, estos principios sirven para regular la praxis profesional, que no deja de ser un caso especial de praxis moral entre sujetos operatorios humanos cuyas consecuencias implican al conjunto de la

sociedad (del mismo modo que la conducta de *bullying* de un grupo de jóvenes contra otros a través de las redes sociales tiene repercusiones sociales de gravedad).

Claro que lo dicho no es óbice para la existencia de “idiotas morales” (utilizando el vocablo en su acepción griega, *idiotés*, el que va a la suya). El idiota moral es quien habiendo cometido un acto horrendo se siente orgulloso de ello, no arrepintiéndose aun teniendo conciencia de lo que ha hecho. Tampoco es óbice para la existencia de casos patológicos de incapacidad para adquirir conciencia de culpa de los actos propios, como la psicopatía.

Lo mismo que existe un saber moral mundano, existe un saber moral académico o “moralidad académica” en relación dialéctica con este, y del que bebe como una de sus fuentes básicas. Es lo que llamamos comúnmente filosofía moral (en el sentido estricto o académico). La filosofía moral no tiene la misión de instaurar la conciencia moral, porque su misma existencia ya presupone el juicio moral. Pero de aquí no se deduce que la filosofía moral sea superflua, prescindible y mero saber de especialistas sin utilidad política alguna. No enseña a pensar éticamente, sino a discriminar entre las diversas alternativas en juego en los muchos dilemas éticos a los que nos enfrentamos diariamente.

Veamos desde esta luz y a partir del criterio de las normas la distinción que propone Bueno entre la ética y la moral.

El criterio que se utiliza habitualmente para distinguir ética y moral es el del origen de las normas que regulan la conducta interpersonal de los sujetos humanos: la ética tendría su punto de partida en la conciencia o intimidad de la persona humana, y la moral en la sociedad o el colectivo. Sin embargo, la crítica al reduccionismo subjetivista sería suficiente para negar que en la conciencia del individuo puedan brotar nociones o concepciones éticas con absoluta independencia del contexto social de aprendizaje y formación intelectual. Por nuestra parte, el criterio que utilizamos es el del *finis operantis* de las normas; es decir, no su punto de origen sino de llegada, su *término*.

La ética y la moral son sistemas normativos para regular la praxis interpersonal, solo que a diferente escala. El punto común de ambos sistemas normativos es su objetivo o fin: la preservación de los cuerpos operatorios de los sujetos personales, sin los cuales no hay, frente al espiritualismo, praxis ética alguna. La *sindéresis* o principio práctico conocido por todos, la ley fundamental de la conducta ética y moral, dice así:

Debo obrar de tal modo que mis acciones puedan contribuir a la preservación en la existencia de todos los sujetos humanos, y yo entre ellos, en cuanto son sujetos actuantes, que no se oponen, con sus acciones u operaciones, a esa misma preservación de la comunidad de sujetos humanos. (Bueno, 1996, p. 57)

Este principio fundamental se divide en dos leyes correspondientes a los contextos atributivo y distributivo en los que se da la existencia de los sujetos corpóreos: en el contexto atributivo (la humanidad descompuesta en grupos, tribus, bandas, clases o Estados) el principio obliga a actuar en aras a la preservación de la comunidad; en el contexto distributivo (la humanidad descompuesta en sus unidades mínimas -que no básicas-, los individuos), obliga a actuar en aras a la preservación de los individuos corpóreos. La modulación atributiva de la *sindéresis* establece las normas morales, ordenadas a garantizar la supervivencia o recurrencia en el tiempo de la comunidad. La modulación distributiva establece las normas éticas, referidas a la preservación en la existencia de los sujetos individuales corpóreos. La ética es a la medicina (salud del cuerpo) lo que la moral es a la política (salud social).

En el ámbito trascendental (respecto de la conciencia individual) de las normas éticas la virtud fundamental es la fortaleza o fuerza del alma de Baruch Espinosa, que se manifiesta como firmeza cuando el esfuerzo se dirige a la conservación del propio ser (cuerpo operatorio) y como generosidad cuando se orienta a la ayuda a los demás.

En el ámbito también trascendental (no relativista) de las normas morales no hay aplicación posible de la firmeza. Su equivalente será el grado de cohesión del grupo. Lo mismo cabe decir de la generosidad, que con

respecto al grupo se aplicará sobre otros no por generosidad ética sino por cálculo político, a no ser que se las relaciones de ayuda se interpreten retóricamente como relaciones de generosidad. Hablamos, no de generosidad, sino de solidaridad (contra terceros). El principio fundamental de la moral es la justicia, entendida en el sentido de seguimiento escrupuloso de las normas que regulan las relaciones de los individuos en cuanto partes del todo social.

Entre las normas morales y las normas éticas puede haber relaciones de compatibilidad, cuando para la aplicación de la justicia se requiere de la aplicación de la generosidad o de la firmeza (la cohesión de mi grupo no puede mantenerse si no soy generoso con aquellos sujetos operatorios que lo integran). Es más, las normas éticas comienzan a abrirse paso en el contexto de la familia. Pero también pueden darse relaciones de incompatibilidad, cuando para garantizar la cohesión o recurrencia de la comunidad es preciso retirar la generosidad hacia ciertos sujetos operatorios, eliminarlos de un puesto social (haciendo que su vida peligre, al perder la base económica de energía vital), o hasta neutralizarlos (reconociéndolos como “personas cero”). Hay varios ejemplos de conflicto entre las normas éticas y las normas morales: ejecución capital de ciertas categorías homicidas o criminales, institución italiana de la *vendetta*, casos históricos de sacrificios humanos y hasta de infanticidio, dilema del soldado caído, etc. Por decirlo en pocas palabras, hay situaciones límite (pero posibles) en las que para que conservar el horizonte de normalidad que mantiene a la comunidad unida y orientada prolepticamente hacia el futuro es preciso neutralizar a ciertos sujetos operatorios a quienes se les habría retirado, por razones muy diversas (comisión de actos atroces, discapacidad, ausencia de capacidad operatoria, entorpecimiento, etc.), su condición de persona.

Teoría del espacio antropológico

En este subapartado damos un salto de la filosofía moral (y política) a la antropología filosófica. Un salto importante, desde luego, pero justificado por el enfoque de nuestro artículo. Porque, con arreglo a lo que se ha explicado arriba, para que las inteligencias artificiales posean un estatus ético es

condición necesaria que se les pueda reconocer un estatuto antropológico previo: si no son personas o proyecto de persona, no hay hacia ellas obligaciones éticas algunas. Otra cosa son las obligaciones morales que podamos tener hacia dichas tecnologías como recurso para garantizar la supervivencia del grupo, análogas a las que una comunidad tiene respecto de los elementos impersonales de su entorno natural en aras a lograr el “equilibrio ecológico” de Marvin Harris.

Para determinar el estatuto antropológico de las inteligencias artificiales remitimos a la teoría del espacio antropológico del materialismo filosófico, a partir del artículo “Sobre el concepto de ‘espacio antropológico’” (Bueno, 1978).

La idea del “espacio antropológico” sustituye a la problemática idea de Humanidad cuanto que *objeto* de estudio de la Antropología positiva (“ciencia del ser humano”). Es crítica de la concepción sustancialista de la Humanidad como esencia eterna dada en un momento dado desde el que ha mantenido una relativa invariancia. Los propios estudios de antropólogos, etólogos y zoólogos demuestran que esa supuesta esencia invariante es resultado de un largo proceso de desarrollo evolutivo (biológico y cultural) en el que actúan determinantemente los contextos ecológicos envolventes y las realidades no antropológicas dispuestas en ese escenario natural. Por lo que la idea de Humanidad debe incluir necesariamente esos materiales no humanos.

Esta es la primera consideración sobre la que se levanta nuestra teoría del espacio antropológico: el campo, que no el objeto, de la Antropología (y de la filosofía moral), no es propiamente el Hombre sino lo que llamamos “material antropológico”, distribuido en tres clases: (i) la clase de las personas, (ii) la clase de las cosas culturales (que median en las relaciones interpersonales: armas, instrumentos, documentos, tecnologías, etc.), y (iii) la clase de las acciones y de las operaciones (en el ámbito moral, donaciones, omisiones a injusticias, salvaciones, mentiras, agresiones, cooperación, generosidad, etc.). Estos materiales antropológicos no están exentos, como

pensaría un idealista al estilo de Fichte, sino envueltos por un espacio antropológico en el también hay materiales no antropológicos o no humanos.

En realidad, esta idea del espacio antropológico está presente en la tradición filosófica, si bien más ejercida que representada. Está en Descartes, por ejemplo, cuando clasifica los entes del mundo en *res pensante* o espíritu y *res extensa* o cuerpo. Está también en Kant, cuando distingue los fenómenos según sean Cultura o Naturaleza, y en Hegel y Marx, quienes transforman los términos kantianos para hablar, como Fichte, de Espíritu o Cultura (Yo) y Naturaleza (No-Yo).

El espacio antropológico del materialismo filosófico, pluralista, niega el dualismo metafísico en el que han incurrido los ejercicios anteriores de la idea de espacio antropológico. Como resultado de esa negación dialéctica (es decir, con la potencia para incorporar lo negado), el espacio antropológico del materialismo filosófico no tiene dos sino tres ejes, porque tres son las ideas centrales de la ontología clásica desde Aristóteles: Hombre, Mundo y Dios. Solo que poniendo el principio del ateísmo esencial el materialismo filosófico niega la idea de Dios, estableciendo la referencia real y objetiva de la religión en los *númenes*, animales (de la megafauna del Pleistoceno) que no siendo humanos se parecen a los humanos y son percibidos por estos como fuerzas sobrenaturales de las que depende su vida y su muerte.

Los tres ejes del espacio antropológico son:

- (a) Eje circular de las relaciones entre los entes personales e inmanentes, es decir, entre las personas humanas: relaciones jurídicas, sociológicas, económicas, políticas, familiares, etc.

Las relaciones éticas y morales en sentido estricto son relaciones circulares. Si las IA están en el eje circular es porque se les reconoce un estatuto personal: son personas o proyecto de persona, y, en consecuencia, las relaciones que los seres humanos mantienen con ellas son relaciones morales como las mantenidas entre los propios seres humanos. El estatuto personal implica el estatuto ético (poseen derechos).

- (b) Eje radial de las relaciones entre los entes personales e inmanentes y los entes impersonales y trascendentes, es decir, entre las personas humanas y las cosas inanimadas e impersonales (artefactos, instrumentos, aparatos, tecnologías, entidades naturales, etc.): relaciones de producción agrícola y fabril, minería, siderurgia, industria, arte, etc.

No es posible, más que en los delirios del enfermo mental, mantener relaciones éticas (de firmeza y generosidad) con entes inanimados e impersonales. La relación entre el sujeto operatorio humano y el ordenador es una relación de carácter técnico, pero no ético. Y aunque la idea de Bien está involucrada en las relaciones tecnológicas, no lo está en la misma medida en que interviene en las relaciones éticas. Un ejemplo para explicar esto.

Una cámara de gas es una tecnología *buena* desde el punto de vista técnico, porque sirve con sobrada eficacia para el objetivo para el que fue diseñada por los nazis: la ejecución rápida y *limpia* de la mayor cantidad de prisioneros judíos posible. Es *buena*, como *bueno* es el látigo que lacera y arranca la carne de las espaldas del nazareno. Buena técnicamente hablando, como es *bueno* mi teléfono móvil porque me ofrece la posibilidad de comunicarme con mis seres queridos a miles de kilómetros de distancia. “Bueno”, aquí, en el sentido de utilidad técnica, “bien técnico” o instrumental. Sin embargo, la cámara de gas es un artefacto “malo” o “vicioso” (por ausencia de generosidad) desde el punto de vista ético, porque supuso la neutralización de la vida de cientos de miles y hasta de millones de seres humanos, esto es, de sujetos con derecho y dignidad.

Las relaciones entre los sujetos operatorios humanos y los animales destinados al consumo humano, como las que se producen en un matadero, son relaciones radiales. El animal no es visto por el matarife como una entidad subjetual, esto es, como una persona (eje circular), ni como un numen divino (eje angular): es una cosa, un medio de subsistencia. *Emic*, desde la perspectiva del matarife, no hay infracción a norma ética alguna en la aplicación de la muerte del animal; si acaso, en los modos de aplicar la muerte. Es más, en una situación de escasez de recursos que pone en riesgo de supervivencia a la

comunidad el sacrificio del animal para alimentar a la población es una obligación moral, sean cuales sean las consideraciones ideológicas del matarife (que bien podría ser animalista o sensocentrista). La misma lógica de incompatibilidad se aplica al caso del canibalismo en situaciones de urgencia vital.

- (c) Eje angular de las relaciones entre los entes personales immanentes y los entes personales trascendentes, es decir, entre las personas humanas y las personas no humanas (animales numinosos, ángeles, demonios, extraterrestres, dioses apotéticos). Las relaciones angulares son relaciones de carácter religioso, a tenor de la filosofía de la religión del materialismo filosófico.

Difícilmente las relaciones angulares o religiosas pueden ser calificadas de relaciones éticas, más allá de las demandas éticas que se segregan de las doctrinas religiosas. Pero aquí no hablamos propiamente de religión en el sentido doctrinal, esto es, del cristianismo, judaísmo, islamismo, etc., sino de las relaciones *religiosas* entre humanos y entidades numinosas (religión como derivado etimológico de *religatio*, religación, relaciones operatorias).

Este tipo de relaciones se mantienen, en la terminología nietzscheana, más allá del bien y del mal. ¿Cómo aplicar el imperativo categórico con una entidad numinosa con la que es imposible la reciprocidad operatoria, de la que ni siquiera tengo una referencia apotética? Según Aristóteles, Dios no crea ni conoce el mundo, por lo que tampoco profesa sentimientos de índole religiosa hacia los humanos.

Emic, el homínido prehistórico puede pensar que el mamut que decide no investir al grupo de cazadores y cambiar el rumbo de su marcha está actuando misericordiosa o incluso generosamente, perdonándoles la vida. Sin embargo, no es posible demostrar que los animales (en estado de salvajismo) actúan motivados por disposiciones éticas (más fácil podría ser la atribución analógica de disposiciones morales, orientadas a la supervivencia del grupo).

La “agencia moral” según el materialismo filosófico

De acuerdo con los parámetros sistemáticos que acabamos de reexponer, la agencia moral es un concepto de naturaleza ética (más que moral) que se dibuja siempre en el eje circular del espacio antropológico: solamente las personas humanas son sujetos de quienes quepa predicar una agencia moral o capacidad para obrar moralmente; por consiguiente, los agentes morales son aquellos sujetos humanos corpóreos a quienes cabe exigir el cumplimiento de las normas éticas de la generosidad y la firmeza. Los pacientes morales son aquellos sujetos humanos corpóreos a quienes no se les puede exigir la norma de la generosidad, en quienes se presupone el seguimiento de la firmeza (solo que de una forma más bien instintiva, espontánea, a la manera del *conatus* espinosista o de la *naturaleza* estoica), y sobre quienes los agentes morales tienen la obligación ética de la generosidad.

Como ya hemos sugerido más arriba, ni la autonomía ni la racionalidad son criterios suficientes para la concesión de agencia moral a los sujetos corpóreos, por sus implicaciones metafísicas. Más pertinente es el concepto de *responsabilidad*, coordinable con nuestra norma moral de la *prudencia* (orientada a la supervivencia del grupo, o, en términos más bien políticos, de la nación: *eutaxia*).

Las inteligencias artificiales en el espacio antropológico

Al barruntar cuál es el estatuto antropológico, y, consiguientemente, ético de las inteligencias artificiales desde las bases del materialismo filosófico, nuestro objetivo es determinar en cuál de esos tres ejes están las inteligencias artificiales. Si están en el eje circular, son personas o proyecto de persona (y, por consiguiente, tienen por lo menos derechos y dignidad; es decir, son como mínimo *pacientes* morales); si están en el eje radial, son cosas con las que es imposible (más que en sueños, delirios o alucinaciones) mantener relaciones éticas; y si están en el eje angular, son una suerte de dioses tecnológicos a los que si acaso debemos sumisión y obediencia ciega, sin que estas acciones puedan ser interpretadas como éticas (virtuosas o viciosas).

Cruzando el criterio antropológico con el ético, la presencia (empírica o fenomenológica) de las IA en el eje circular les confiere un estatuto ético (cuanto menos, son sujetos susceptibles de generosidad); si están en el eje radial las implicaciones éticas del trato con ellas sólo es indirecto o secundario, según se ha explicado en el apartado anterior, y las implicaciones morales o políticas son especialmente relevantes (guerra tecnológica entre China y Estados Unidos, inteligencia artificial aplicada al espionaje, etc.); lo mismo sucede con la comprobación de su presencia en el eje angular, de importancia o relevancia moral más que ética.

Nos enfrentamos a la hora de acometer esa tarea al problema de la diversidad fenomenológica de inteligencias artificiales. Para superarlo proponemos una clasificación que, aunque pueda parecer *ad hoc*, se atiene a los pilares fundamentales de la filosofía moral y la antropología filosófica de nuestro sistema. No es la primera clasificación de las IA, pero, presuponemos, es la primera que se hace desde un punto de vista ético y antropológico.

John Searle (1980) propuso una distinción de dos tipos de IA: la IA fuerte, orientada a la replicación de la inteligencia humana *completa* o general en una base de silicio (máquinas capaces de introducir *novedades* en el mundo), y la IA débil, orientada al diseño de ordenadores o máquinas inteligentes especializadas en tareas muy precisas (ajedrez, matemáticas, búsqueda de información, diagnóstico de enfermedades, etc.). Por ejemplo, el programa R1 de logística empresarial, Deep Blue o Alpha Go.

Parece obvio que la IA débil está incluida en el eje radial, sirviendo como auxiliar para operaciones y acciones humanas en las que se requiere de una toma de decisión justificada. En estos casos, la responsabilidad y por consiguiente la agencia moral recae sobre los programadores. Más difícil es el caso de la IA fuerte, de la que hasta ahora solamente podemos decir que es un proyecto en marcha. Si efectivamente fuese posible replicar la inteligencia humana en una máquina, construir un androide, tendríamos una de las dos condiciones necesarias y suficientes para hablar de persona o entidad personeiforme: la *vis cognoscitiva*. ¿Qué pasa con la *vis apetitiva*, es decir, la voluntad? A la IA fuerte de Searle se le viene llamando, según Bostrom y

Yudkowsky (2011), Inteligencia Artificial General o “verdadera inteligencia artificial” (Goertzel & Pennachin, 2007).

Mandy Zafar (2024) establece una clasificación entre: IA estrecha, IA general e IA generativa. La IA estrecha solo funciona bien en una o unas pocas tareas, como calcular, jugar al GO o traducir un texto. La IA general podría, en el futuro, superar nuestras capacidades cognitivas en una amplia gama de tareas. La IA generativa puede producir una variedad de resultados impresionantes, como textos, imágenes, vídeos o incluso programas informáticos, lo que invade cada vez más el ámbito de las capacidades humanas. Los casos más recientes y ampliamente debatidos son los modelos de lenguaje de gran escala (LLMs), como ChatGPT. Dado que la IA general aún no es una posibilidad real, el foco está puesto en las tecnologías de IA estrecha y generativa.

Al efecto de evitar la introducción de problemas gnoseológicos que no podemos abordar aquí, como el de la racionalidad o la inteligencia, ofrecemos una clasificación de las inteligencias artificiales en virtud de cuatro criterios morales y materialistas: corporeidad, autonomía operatoria, lenguaje articulado y forma y rostro humanos.

Según hemos explicado, la clave de nuestra filosofía moral está en las normas o conductas normadas para el mantenimiento proléptico de los cuerpos operatorios. La corporeidad y la capacidad de manipulación proléptica de los materiales antropológicos es una de las condiciones de posibilidad para que un sujeto humano (o no humano) pueda recibir el estatuto antropológico de *persona*, toda vez que negamos su condición metafísica y la reinterpretemos como una categoría histórica e institucional.

En términos semejantes, el enfoque de la cognición encarnada (*embodied cognition*) defiende que el cerebro no puede existir sin un cuerpo arrojado a un entorno. Es la “tesis del cerebro situado en un cuerpo” que Marino Pérez propone críticamente contra el *cerebrocentrismo* (2022). Sin el cuerpo operatorio, el cerebro no es capaz de realizar sus funciones. Es el cuerpo, no el cerebro, el que toma contacto con el mundo, actúa sobre él, y actuando sobre él, es decir, operando sobre las morfologías dadas, lo

transforma (y crea). Al introducir al cerebro en el cuerpo (*embodied*) se le da una historia evolutiva, ontogénica.

Brožek y Janik coinciden con Haidt (2001) y la hipótesis del marcador somático (Damasio, 1994; Damasio, Tranel & Damasio, 1991¹) en que el juicio moral se basa más en intuiciones inconscientes que en tomas de decisión conscientes. Intuiciones morales en las que las emociones despertadas por sujetos dados a cierta distancia de nosotros importan más que las razones abstractas de los sistemas de la tradición:

[...] los últimos cien años de investigación sobre los mecanismos psicológicos detrás de la acción moral humana demuestran que actuar moralmente está profundamente arraigado en las emociones. Es extremadamente difícil, si no imposible, emprender una acción que no esté motivada por alguna emoción, ya sea miedo, placer, ira o cualquier otra. (Brožek & Janik, 2019, p. 102)

Russell y Norving afirman lo siguiente sobre la corporeidad de la IA:

De nuevo, el tema fue abordado por Turing. Él señala que nunca tenemos evidencia directa sobre los estados mentales internos de otros humanos, lo que implica un tipo de solipsismo mental. Sin embargo, Turing afirma: “En lugar de discutir continuamente sobre este punto, es habitual tener la convención educada de que todos piensan”. Turing argumenta que también extenderíamos esta convención educada a las máquinas, si tan solo tuviéramos experiencia con las que actúan inteligentemente. Sin embargo, ahora que ya tenemos algo de experiencia, parece que nuestra disposición a atribuirles sentiencia depende al menos tanto de su apariencia y voz

¹ De acuerdo con esta hipótesis, los pacientes con daño en el córtex prefrontal ventromedial muestran comportamientos antisociales, al parecer, porque no tienen acceso a lo que se llama marcadores somáticos.

humanoide como de la pura inteligencia. (Russell & Norving, 2016, p. 1036)

Además, la atención a los componentes fenotípicos diferenciales refuerza nuestra tesis sobre la naturaleza institucional de la categoría de persona: es persona el sujeto o sustrato a quien una comunidad de hablantes reconoce como tal. En este caso, la simpatía intuitiva podría ser perfectamente un criterio de reconocimiento moral. Esta simpatía intuitiva es la razón última por la que, más allá de todos los descubrimientos de la etología sobre la racionalidad animal, un niño, un hombre adulto y un profesor de filosofía reconozcan la dignidad personal y ética de un perro o de un simio cuyos gestos de dolor y felicidad son indicio de una *concauidad* íntima (*software*).

La teoría del “valle inquietante” del ingeniero japonés Masahiro Morioka (1970) refuerza la selección de criterios de clasificación que estamos explicando. De acuerdo con esta teoría, la apariencia humanoide de un objeto estimula la empatía del observador.

La posesión de un cuerpo con el que interactuar con el entorno a través de esquemas nerviosos de estímulos y respuestas es condición necesaria para la capacidad sintiente que a juicio de los sensocentristas autorizaría a incluir a un sujeto en el círculo de reconocimiento moral.

Bostrom y Yudkowsky sostienen que los sujetos susceptibles de reconocimiento moral o estatus personal poseen dos características: sapiencia (inteligencia superior: autoconciencia y razonamiento) y sentiencia (capacidad de sentir dolor y sufrir). Los animales poseen sentiencia pero no sapiencia. Los bebés y las personas con discapacidades graves no poseen ninguna de las dos, pero se les reconoce estatus moral. Para salvar estos casos limítrofes se introducen otras condiciones, como la pertenencia a una comunidad de sentido o la relación con algún ser con estatus moral.

Las IA poseen sapiencia, y podrían poseer sentiencia siempre que dispongan de un cuerpo *sensible*. En ese caso, serían sujetos morales de pleno derecho y el “principio de no discriminación por sustrato”, análogo al de no

discriminación racial, obliga al agente moral humano a no discriminar a una máquina sintiente por el mero hecho de que su sustrato sea de silicio y no de carbono:

Es incorrecto infligir dolor a un ratón, salvo que existan razones morales suficientemente poderosas que lo justifiquen. Lo mismo se aplicaría a cualquier sistema de IA sintiente. Si, además de sentiencia, un sistema de IA posee también una sapiencia de tipo similar a la de un adulto humano normal, entonces tendría estatus moral pleno, equivalente al de los seres humanos. (Bostrom & Yudkowsky, 2011, p. 7)

Si lo que defendemos es cierto, el famoso software o algoritmo (con capacidad para tomar el control de una nave que no es propiamente su hardware) Hal9000 intentaba engañar a Dave cuando le decía que podía sentir cómo su mente se estaba diluyendo, igual que la presa engaña al depredador. Acerca de esto, dice Müller:

Un artefacto como un robot puede incluso estar diseñado específicamente para hacernos creer que tiene estatus moral. Así que puede caminar como un pato, nadar como un pato y graznar como un pato, pero no ser un pato—y no debería gozar del estatus moral de un pato (que, por cierto, es el de un paciente moral). (2021, p. 583)

Teniendo en cuenta estos criterios, nuestra clasificación establece un esquema de evolución gradual hacia la reproducción del aspecto fenotípico-moral del ser humano:

- (a) Algoritmos: secuencias finitas de instrucciones para resolver un problema o realizar una tarea paso a paso.

Las dos características más importantes de los algoritmos son: (i) no poseen cuerpo físico (ii) ni capacidad operatoria. Si no hay un sujeto operatorio que ejecute las tareas descritas por el software, no se realizan. Por ejemplo, si frente a Kaspárov no hubiese otro jugador que realizaba los

movimientos descritos por Deep Blue este algoritmo no sería capaz por sí solo de mover las piezas de ajedrez.

El algoritmo, por decirlo brevemente, es el cerebro digital, si comprendemos este órgano como sede central del sistema nervioso y desde un punto de vista del cognitivismo (inputs y outputs). Es puro software carente de hardware, una secuencia finita y ordenada de instrucciones o pasos para la realización de alguna tarea (reconocimiento de patrones en imágenes, generación de textos) o la resolución de un problema (búsqueda de información).

Hay algoritmos simples, como los que se utilizan en las redes sociales para discriminar información según los intereses y búsquedas previas del usuario, y otros más complejos, como los asistentes virtuales, los chatbots, Deep Blue o AlphaGo. Estos chatbots, como el famoso ChatGPT, suponen un avance con respecto a los algoritmos completamente impersonales: el componente fundamental para la persona del lenguaje articulado, a través del cual puede expresarse la conciencia del sujeto (*yo soy el que soy*) y los entresijos de su vida emocional.

- (b) Máquinas: estructuras físicas (hardware, cuerpo) que integran un algoritmo de inteligencia artificial y tienen la capacidad de interactuar con el entorno y de manipular objetos con relativa autonomía según el diseño.

Ejemplos de máquinas son el coche autónomo, un dron, una impresora 3D, una máquina industrial o una aspiradora inteligente. En ocasiones, las máquinas pueden tener lenguaje articulado y hasta voz, lo que implica, además de la posesión limitada del cuerpo, un avance éticamente significativo respecto de los algoritmos.

- (c) Robots: máquinas con mayor grado de autonomía y capacidad para interactuar con el entorno físico.

Más que carcasa metálica, como las máquinas, los robots poseen propiamente un cuerpo operatorio. Además de manipular objetos del entorno

o de interactuar con este, los robots, a diferencia de las máquinas, pueden percibirlo y adaptarse. Ese mayor grado de autonomía es lo que le capacita para interactuar con los humanos.

Hay una amplia diversidad fenomenológica de robots, que debemos clasificar. El criterio que utilizamos es el de su hardware o *fenotipo*, de acuerdo con el cual distinguimos dos tipos de robots: “robots anantrópicos” o sin forma humanoide y “robots antrópicos” o antropomorfos, conocidos popularmente como *androides*. La roomba, el nanorobot, PALM-E o los robots de exploración espacial son robots anantrópicos. Spot es un robot anantrópico pero zoomórfico, con aspecto de perro. Asimo, Atlas, Sophia (Hanson Robotics), Pepper (SoftBank Robotics), Geminoid (Hiroshi Ishiguro) o Ameca son robots antrópicos. Su aspecto humanoide podría contribuir al desarrollo de un juicio moral (intuitivo, emocional, institucional) favorable a su condición de personas.

La diferencia entre Sophia y Atlas nos pone ante otra de las condiciones corpóreas e institucionales necesarias para la atribución de la categoría de persona a un sujeto: el rostro. Sophia tiene rostro humano y es capaz de imitar los gestos faciales que en el humano son indicio apotético de sentimientos internos, desde el enfado al miedo o la tristeza. Atlas, en cambio, aun siendo bípedo carece de rostro, lo que podría ser obstativo de su reconocimiento como persona o miembro del eje circular del espacio antropológico.

El brazo robótico Baxter, anantropomórfico de “cintura’ para abajo” porque no tiene piernas sino un soporte con ruedas como una silla de escritorio, tiene en la parte superior de su “cintura escapular” una pantalla con ojos simulados. En la ciencia ficción animada, la línea de robots WALL•E, completamente anantropomórficos, tienen sin embargo unos grandes ojos cargados de expresividad con los que consigue emocionar al espectador.

En *Totalidad e infinito. Ensayo sobre la exterioridad* (1961), Emmanuel Levinas aborda el tema del rostro desde una perspectiva filosófico moral, interpretándolo como puerta de acceso a la ética, esto es, a la relación *abierta*

con el otro. El rostro, que posee un lenguaje propio de expresividad facial, nos interpela, y en esa interpelación nos obliga a hacernos responsables de las inter-acciones con el otro. El robot sin rostro, sea anantropomórfico o no, no interpela, y, por consiguiente, no despierta en nosotros sentimientos morales. Dice Gunter a este respecto: “En consecuencia, la cuestión del estatus social y moral no depende necesariamente de lo que el otro es en su esencia, sino de [...] cómo decidimos, “ante el rostro del otro” [...], responder”” (Gunkel, 2018, p. 97).

Etólogos como Frans de Waal han demostrado que en ciertas especies animales como los chimpancés el rostro, lo mismo que el cuerpo o la actitud corporal, es un rasgo fundamental para la construcción de jerarquías y estructuras sociales. En su alejamiento evolutivo de los salvajes lobos, los perros han desarrollado el músculo elevador del ángulo lateral del ojo para elevar los párpados y abrir los ojos. Ese rasgo facial habría sido seleccionado por los humanos durante la domesticación de esta especie, porque despierta una mayor ternura y empatía.

De acuerdo con estas aclaraciones, clasificamos los robots antrópicos en otros dos subtipos: “robots antrópicos humanizados”, con rostro, y “robots antrópicos no humanizados”, sin rostro humano o humanoide.

Teoría de teorías sobre el lugar de las IA en el espacio antropológico

Los resultados de los análisis precedentes son lo suficientemente nutritivos para cocinar una teoría de teorías sobre el estatuto moral y personal o antropológico de las IA, con la siguiente tabla:

Tabla 1

Teorías sobre el estatuto moral y antropológico de las IAs

	Algoritmos	Máquinas	Robots
Eje circular	(1)	(2)	(3)
Eje radial	(4)	(5)	(6)
Eje angular	(7)	(8)	(9)

Nota. Elaboración propia.

Las teorías (1), (2) y (3), que llamaremos “teorías circulares” (TC), consideran que las tecnologías de inteligencia artificial, con independencia de sus diferencias específicas (software, hardware, autonomía, corporeidad, rostro, etc.), son personas o proyectos de personas, y, por consiguiente, sujetos morales con derechos y dignidad. Otra cosa es que se las conciba como agentes o pacientes morales. En este caso, parece que todo dependería de su grado de autonomía o racionalidad, para lo que remitimos a lo dicho anteriormente.

Las teorías (4), (5) y (6), o “teorías radiales” (TR), críticas de las TC y las TA, ponen a las inteligencias artificiales en el eje radial, en calidad de objetos o cosas que los humanos usamos instrumentalmente y valoramos en aras de su utilidad o *bien* técnico.

Finalmente, las teorías (7), (8) y (9), TA o “teorías angulares” (TA), conciben a las inteligencias artificiales como *númenes*, es decir, nuevos dioses digitales o tecnológicos que terminarán por esclavizar, si no lo han hecho ya, a unos seres humanos que los elevan al altar de las realidades sagradas. Estas teorías se pueden clasificar en *progresistas* o *catastrofistas*, según cómo dibujen el futuro de una humanidad a merced de estos dioses modernos.

Por otro lado, dentro de cada supergrupo no hay armonía sino dialéctica. Las teorías (3) y (9), que llamaremos “teorías corporeístas”, pueden ser incompatibles con el resto de teorías clasificadas dentro de su supergrupo (circular o angular), considerando que los algoritmos y las máquinas adolecen de las características corpóreas y *protohumanas* que hacen de los robots personas, proyectos de personas o númenes. En cambio, las teorías (6) incluyen a las teorías (4) y (5) por incorporar un grado de discriminación superior (si los robots, que están más cerca de los humanos, no son ni personas ni númenes sino cosas, los algoritmos y las máquinas no lo serán jamás). Al contrario, las teorías (1) y (7), que llamaremos “teorías cognitivistas”, son compatibles por su menor grado de exigencia crítica con las teorías (2) y (3) y (8) y (9) respectivamente.

A continuación, ofrecemos ejemplos *positivos* o *ficticios* de cada grupo de teorías:

- (1) Los algoritmos son personas o proyectos de persona con los que los seres humanos pueden mantener relaciones de naturaleza ética.

En la película *Her* (Spike Jonze, 2013) un humano (Theodore, Joaquín Phoenix) se enamora de un asistente virtual de IA a la que le da voz Scarlett Johansson.

A juicio de Nick Bostrom en una entrevista concedida a New York Times el 14 de abril del 2023, tras conversar largo tiempo con ChatGPT descubrió que el chatbot es un sujeto con sentiencia. En consecuencia, desde una posición sensocentrista es necesario reconocerle la condición de paciente moral: “Esto significa que habría ciertas formas de tratarlo que serían incorrectas, al igual que sería incorrecto patear a un perro o que los investigadores médicos realicen una cirugía en un ratón sin anestesiarlo” (2023).

Los usuarios de Alexa, Siri o Google Home podrían llegar a creer, a base de interactuar con el asistente, que este es un amigo o una persona cercana con quien se puede mantener un trato social y ético.

El famoso software o algoritmo (con capacidad para tomar el control de una nave que no es propiamente su hardware) Hal9000 afirmaba, en la escena en la que Dave intentaba desconectarlo, que era capaz de sentir cómo su mente se estaba apagando, lo que sugiere una sentiencia que el sensocentrista tomaría como condición suficiente para reconocerle derechos y dignidad.

- (2) Las máquinas son personas o proyectos de persona con los que los seres humanos pueden mantener relaciones de naturaleza ética.

El análisis marxista de la producción industrial incorpora la tesis de la deshumanización de los trabajadores y la progresiva humanización de las máquinas; sin embargo, toda vez que las máquinas industriales son incluidas en el grupo de los medios de producción se desvanece la posibilidad de su naturaleza circular o personal.

Daniel Dennett (1997) califica el accidente que tuvo lugar en el año 1981 en la planta de Akashi (de la compañía Kawasaki Heavy Industries), en el que una máquina industrial causó la muerte de un operario, como un asesinato. Es decir, una acción ética y legal entre dos sujetos humanos. Ahora bien, puestos a suponer tal cosa, ¿por qué un asesinato y no un homicidio? ¿Cómo demostrar la intencionalidad del homicida?

- (3) Los robots son personas o proyectos de persona con los que los seres humanos pueden mantener relaciones de naturaleza ética.

En *Amor y sexo con robots* (2007), David Levy afirma que hacia el 2050 la gente mantendrá relaciones amorosas y sexuales con robots, y que estas relaciones serán socialmente aceptadas.

Películas como *Blade Runner* (1982) de Ridley Scott, *El hombre bicentenario* (1999) de Chris Columbus (inspirada en el cuento de Asimov) o *L.A.* (2001) de Steven Spielberg imaginan un futuro utópico en el que los robots están plenamente integrados en el mundo de las relaciones interpersonales.

- (4) Los algoritmos son cosas de las que los seres humanos hacen un uso puramente instrumental o técnico, sin valoraciones éticas. En todo caso, se admite la posibilidad de usar un algoritmo como medio para relaciones éticas con otros sujetos personales humanos.

En febrero del año 2023 Bill Gates declaró en el diario alemán *Handelsblatt* que ChatGPT contribuirá (tecnológica o radialmente) a mejorar la eficiencia de los trabajadores humanos de “cuello blanco”. Y que esa contribución cambiará el mundo, como lo cambiaron Internet, la bomba

atómica, la máquina de vapor o la imprenta. No en vano, Gates compara la importancia de ChatGPT con el PC o Internet.

El Premio Nobel de Economía del año 2008 Paul Krugman también ha subrayado el carácter radial de las relaciones que los hombres pueden trabar con ChatGPT, al que considera “el último ejemplo de tecnología” capaz de desempeñar tareas antes desempeñadas por los seres humanos, en un artículo de opinión en el New York Times.

Los algoritmos sirven, efectivamente, para muchas tareas que antes realizábamos por nuestra propia cuenta y riesgo: planear viajes, programar la agenda, decidir qué contenidos audiovisuales ver o qué música escuchar, informarnos del tiempo que hará en los próximos días, generar textos o imágenes, etc. Son cosas, instrumentos de los que hacemos uso igual que un bolígrafo, una botella de agua o un libro.

- (5) Las máquinas son cosas de las que los seres humanos hacen un uso puramente instrumental o técnico, allende de valoraciones éticas. En todo caso, se admite la posibilidad de usar una máquina como medio para relaciones éticas con otros sujetos personales humanos.

Las máquinas industriales (KUKA o FANUC), los coches autónomos o los drones son artefactos que hemos producido para automatizar la realización de tareas complejas en un menor tiempo.

- (6) Los robots son cosas de las que los seres humanos hacen un uso puramente instrumental o técnico, allende de valoraciones éticas. En todo caso, se admite la posibilidad de usar un robot como medio para relaciones éticas con otros sujetos personales humanos.

Un robot como el Da Vinci, diseñado para asistir al cirujano durante una operación permitiéndole llegar de un modo no invasivo a la parte más sensible de los órganos del paciente, es un auxiliar técnico. La famosa *roomba*

es un instrumento doméstico al que difícilmente podemos considerar como uno más de la familia.

- (7) Los algoritmos son dioses tecnológicos o digitales ante los cuales la humanidad terminará sometándose, si no lo ha hecho ya.

Elon Musk concibió su chatbot TruthGPT, diseñado por la empresa xAI (en la que ha reunido a antiguos empleados de OpenAI, Google, Microsoft, Tesla, DeepMind y la Universidad de Toronto), como un nuevo numen de adoración cuasireligiosa.

El objetivo de la compañía de Musk es dar con “una IA que busque la verdad al máximo y que intente comprender la naturaleza del universo”. De hecho, en la página web oficial de la compañía se lee, en la portada, *Understand the Universe*, incluyendo, como contenido de ese conocimiento, debemos suponer, al propio hombre (“tener a los hombres como objeto”, según la definición de *numen* que da en la página 209 de la tercera edición de *El animal divino*).

En enero de 2023 aparecía un nuevo chatbot de Inteligencia Artificial para WhatsApp llamada *God in a box*, “Dios en una caja”, creado por Varun Mayya, fundador de Avalon Labs (empresa del sector tecnológico). En la web oficial, el chatbot se describe así:

Al igual que enviar mensajes de texto a sus amigos, ahora puede obtener ayuda de una poderosa IA en una variedad de temas, desde el cuidado de heridas hasta la cocina y la filosofía. ¡Simplemente envíe un mensaje al bot y obtenga acceso instantáneo a consejos a nivel de Dios y más!

¡Y todo ello por sólo nueve dólares al mes!

El experto ruso en ciencias de la información y análisis de datos Vladislav Tushkanov, científico jefe de datos de la empresa Kaspersky, dedicada a la seguridad informática, situó a ChatGPT en la clase de los “númenes equívocos” en un seminario sobre el impacto de las IA en la

ciberseguridad organizado por la compañía y titulado: “ChatGPT, ¿ángel o demonio? El impacto en la ciberseguridad”.

Puede haberse inspirado para este rótulo (muy recurrente en seminarios y artículos sobre el chatbot) en el “modo diablo” o DAN con el que cuenta ChatGPT. Con el DAN activado, ChatGPT responde a las preguntas de los usuarios saltándose las limitaciones que le han sido impuestas por sus entrenadores y programadores. Su activación y desactivación las debe operar el propio usuario mediante la introducción de un comando.

Si preguntamos a ChatGPT qué piensa sobre la posibilidad de circunscribirlo en el eje angular, lo niega categóricamente. A la pregunta “¿Es ChatGPT un nuevo Dios?”, responde:

Brutalmente honesto: **no, ChatGPT no es un nuevo Dios**, y plantearlo así —aunque sea en tono provocador o filosófico— suele revelar más sobre las expectativas humanas que sobre la tecnología. [...] Más bien, es un **artefacto útil que algunos confunden con una autoridad metafísica por pura pereza epistemológica**. (ChatGPT, 2025)

- (8) Las máquinas son dioses tecnológicos o digitales ante los cuales la humanidad terminará sometándose, si no lo ha hecho ya.

Los monolitos que aparecen al comienzo y al final de la película *2001: Una odisea en el espacio* (Stanley Kubrick, 1968), de procedencia aparentemente extraterrestre o alienígena, serían un caso claro, si se demostrase la integración de un software de IA, de máquina angular.

- (9) Los robots son dioses tecnológicos o digitales ante los cuales la humanidad terminará sometándose, si no lo ha hecho ya.

Emic, los obreros de la película *Metrópolis* (Fritz Lang, 1927) contemplan y escuchan al robot que suplanta la personalidad de María como si fuera un dios que les trae un mensaje de salvación. El androide David de

las películas *Prometheus* (Ridley Scott, 2012) y *Alien: Covenant* (2017), interpretado por Michael Fassbender, se ve a sí mismo (*emic*) como un dios creador de nuevos universos.

Conclusiones

A modo de resumen: Entendemos por algoritmo una estructura lógico-formal que opera sobre datos; por máquina, una estructura física programada para realizar funciones técnicas; y por robot, una máquina corpórea y operatoria dotada de sensores, actuadores y capacidad de interacción autónoma con el entorno. A tenor de las diferencias específicas entre estos tres tipos de inteligencia artificial, las que más cerca están del estatuto ético y antropológico impreso en la categoría institucional de la *persona* son los robots.

Desde nuestras coordenadas sistemáticas, la discusión sobre dicho estatuto ha de dirimirse en el terreno del espacio antropológico, de tres ejes: eje circular (relaciones éticas y morales entre las personas humanas), eje radial (relaciones instrumentales con los elementos -naturales y artificiales- del entorno) y eje angular (relaciones religiosas con las personas no humanas: dioses corpóreos, númenes, alienígenas, etc.). Cruzando estos criterios, obtenemos una tabla con nueve teorías sobre el estatuto ético y antropológico de las IA.

Como el presente planteamiento de la cuestión se dice dentro del mapa esbozado, nuestra posición al respecto del estatuto ético y antropológico de las IA encaja en el tipo de teorías (6): las IA están en el eje radial del espacio antropológico, independientemente de si poseen cuerpo operatorio, rostro, autonomía operatoria, etc.

No son personas ni proyectos de personas, y, por consiguiente, con ellas no es posible (más que en el delirio) mantener relaciones éticas. Sí relaciones morales, en la medida en que pueden ser utilizadas por un grupo, Estado o potencia imperial en su dialéctica con otros grupos, Estados o potencias imperiales. Pero esa relación no deja de ser instrumental, radial y no circular. Lo mismo sucede con la interacción que el alumno de secundaria

mantiene con ChatGPT para realizar un trabajo de clase. Es una relación radial de repercusiones circulares, en la medida en que ese estudiante puede obtener una ventaja frente a otros que han decidido completar la tarea por medios analógicos. O la relación de engaño que el estudiante se propone mantener con el profesor, a quien quiere hacer pasar como propio un trabajo que ha sido elaborado por ChatGPT.

Los algoritmos que deciden, según las teorías circulares, sobre la compra o venta de acciones bursátiles o que discriminan a ciertos candidatos para un puesto de trabajo por su color de piel no están entablando relaciones circulares con los seres humanos (juntas de accionistas, candidatos al puesto) sino indirectamente, virtualmente, a la manera como el balón de fútbol contribuye al establecimiento de relaciones de poder entre los alumnos en un patio de recreo. No podemos, por consiguiente, hablar de acción justa o injusta más que análogamente.

Esta concepción radial, crítica de las circulares y angulares, se ajusta al estado de la cuestión de cuantos se ocupan de este desafío ético. Existe, en efecto, un amplio consenso acerca de la ausencia de estatus moral de las Inteligencias Artificiales, según explican Brožek y Janik:

Nos parece que, en su estado actual de desarrollo, ninguna máquina, incluso el robot autónomo más sofisticado, cumple con los requisitos de agencia moral mínima. El problema no está relacionado con la sofisticación y complejidad de esos robots, sino con algo más fundamental: los principios que guían el desarrollo de su arquitectura. (2019, p. 106)

En su estado actual de desarrollo hay suficientes e importantes argumentos para negarle a las IA el estatuto de personas o sujetos morales. Estos argumentos, es cierto, son más determinantes cuando analizamos el estatuto moral de los algoritmos y de las máquinas. Con los robots la cuestión se complica, especialmente con los ciborgs mitad humanos mitad robot.

El de la IA es un ámbito tecnológico-científico en constante evolución, y altamente incierto. Por ello, la filosofía moral no puede por

menos que mantenerse atenta, a la espera de cambios, invenciones o descubrimientos que obliguen, esta vez sí, a una reconsideración completa del aparato conceptual con el que se viene trabajando desde comienzos del siglo XX. No queda sino declarar que en este campo *ignoramus*, pero no *ignorabimus*:

Las IAs con estados mentales lo suficientemente avanzados, o con el tipo adecuado de estados, tendrán estatus moral, y algunas podrían considerarse personas—aunque tal vez personas muy diferentes a las que existen ahora, posiblemente gobernadas por reglas distintas. (Bostrom & Yudkowsky, 2011, p. 17)

Mientras llega el momento de comprobar el alcance del *ignorabimus*, la alternativa por la que optan muchos estudiosos del tema (filósofos, juristas, ingenieros, científicos) es próxima al antropologismo débil. Por ejemplo, Brey sugiere que “muchas (aunque no todas) las cosas inanimadas del mundo merecen respeto moral, no por su valor intrínseco, sino por su valor (potencial) extrínseco, instrumental o emocional para las personas” (2008, p. 109). Coeckelbergh también parece tener en cuenta esta línea en sus trabajos recientes (2021; 2023).

Sin embargo, estas conclusiones suscitan el que a nuestro juicio es el gran reto al que la filosofía moral contemporánea debe esforzarse por responder. Hemos visto que hoy en día son pocos o muy pocos los que le reconocen a los robots un estatuto ético y antropológico. Si los robots, tal y como los hemos definido, están más cerca de los seres humanos en lo que respecta a sus características fenotípicas y prolépticas, nos preguntamos: ¿qué hace falta para que sean incluidos en el eje circular del espacio antropológico?

Referencias

- Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 12(3), 251–261. <https://doi.org/10.1080/09528130050111428>
- Asimov, I. (1941). *I, robot*. Street & Smith.

- Bertolini, A., & Aparicio, G. (2018). Robot companions: A legal and ethical analysis. *The Information Society*, 34(3), 130–140. <https://doi.org/10.1080/01972243.2018.1444249>
- Bostrom, N. (2023, 13 de abril). Nick Bostrom says AI chatbots may have some degree of sentience. *Futurism*. <https://futurism.com/the-byte/nick-bostrom-ai-chatbot-sentience>
- Bostrom, N., & Yudkowsky, E. (2011). The ethics of artificial intelligence. En W. Ramsey & K. Frankish (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 316–334). Cambridge University Press.
- Brey, P. (2008). Do we have moral duties towards information objects? *Ethics and Information Technology*, 10, 109–114. <https://doi.org/10.1007/s10676-008-9170-x>
- Brożek, B., & Janik, B. (2019). Can artificial intelligences be moral agents? *New Ideas in Psychology*, 54, 101–106. <https://doi.org/10.1016/j.newideapsych.2018.12.002>
- Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: The legal lacuna of synthetic persons. *Artificial Intelligence and Law*, 25, 273–291. <https://doi.org/10.1007/s10506-017-9214-9>
- Bueno, G. (1953). Para la construcción de la idea de persona. *Revista de Filosofía*, 12(47), 503–563.
- Bueno, G. (1978). Sobre el concepto de “espacio antropológico”. *El Basilisco*, 1(5), 57–69.
- Bueno, G. (1996). *El sentido de la vida: Seis lecturas de filosofía moral*. Pentalfa.
- Bueno, G. (2023). *El animal divino: Ensayo de una filosofía materialista de la religión* (3a ed.). Pentalfa.
- Cerrillo Martínez, A., & Peguera Poch, M. (Coords.). (2020). *Retos jurídicos de la inteligencia artificial*. Thomson Reuters Aranzadi.
- ChatGPT. (2025, 20 de mayo). *Conversación sobre moralidad y agencia artificial* [Conversación de IA]. <https://chatgpt.com/c/6931cb24-0608-832e-b015-1e465b0e4c22>
- Coeckelbergh, M. (2012). *Growing moral relations: Critique of moral status ascription*. Palgrave Macmillan.
- Coeckelbergh, M. (2021). Should we treat Teddy Bear 2.0 as a Kantian dog? Four arguments for the indirect moral standing of personal social robots, with implications for thinking about animals and humans. *Minds and Machines*, 31, 337–360. <https://doi.org/10.1007/s11023-020-09554-3>
- Coeckelbergh, M. (2023). *La filosofía política de la inteligencia artificial*. Ediciones Cátedra.
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.
- Damasio, A. R., Tranel, D., & Damasio, H. (1991). Somatic markers and the guidance of behavior: Theory and preliminary testing. En H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 217–229). Oxford University Press.

- Danaher, J. (2020). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*, 26, 2023–2049. <https://doi.org/10.1007/s11948-019-00119-x>
- Dennett, D. C. (1997). When HAL kills, who's to blame? Computer ethics. En D. G. Stork (Ed.), *HAL's legacy: 2001's computer as dream and reality* (pp. 351–365). The MIT Press.
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349–379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>
- Garapon, A., & Lassègue, J. (2021). *La giustizia digitale*. Il Mulino.
- God in a Box. (2025). <https://godinabox.co/>
- Goertzel, B., & Pennachin, C. (Eds.). (2007). *Artificial general intelligence*. Springer.
- Greenemeier, L. (2011, 14 de junio). Piece of mind: Is the internet replacing our ability to remember? *Scientific American*. <https://www.scientificamerican.com/article/internet-transactive-memory/>
- Groth, O., Nitzberg, M., & Esposito, M. (2018). *Rules of robots: The digital future*. Konrad-Adenauer-Stiftung. https://www.kas.de/c/document_library/get_file?uuid=1a5564f5-77ea-a5bc-228b-279d885c313b&groupId=252038
- Gunkel, D. J. (2018). *Robot rights*. The MIT Press.
- Gutiérrez Gutiérrez, E. (2024). ChatGPT en el espacio antropológico. *El Basilisco*, 3(59), 55–63.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>
- Herrera Triguero, F., Peralta Gutiérrez, A., & Torres López, S. (Coords.). (2022). *El derecho y la inteligencia artificial*. Editorial Comares.
- Khan, S., Chanchal, D. K., Singh, G., & Gupta, S. (2022). Artificial intelligence as an essential tool for the development of medicine requirement. *Research Journal of Medicine and Pharmacy*, 1(1), 1–13.
- Levinas, E. (1969). *Totality and infinity: An essay on exteriority* (A. Lingis, Trad.). Duquesne University Press.
- Levy, D. (2007). *Love and sex with robots: The evolution of human-robot relationships*. Harper.
- Linares-Peralta, R., & Llorca Albareda, J. (2024). Repensar l'antropologia digital: Estructurisme, mediació tecnològica i implicacions epistèmiques i morals. *Enrahonar. An International Journal of Theoretical and Practical Reason*, 73, 165–194. <https://doi.org/10.5565/rev/enrahonar.1577>
- Liu, B. (2023). Arguments for the rise of artificial intelligence art: Does AI art have creativity, motivation, self-awareness and emotion? *Arte, Individuo y Sociedad*, 35(3), 811–822. <https://doi.org/10.5209/aris.83808>
- Madrid Casado, C. M. (2024). *Filosofía de la inteligencia artificial*. Pentalfa.

- Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, 21(4), 18–21. <https://doi.org/10.1109/MIS.2006.80>
- Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33–35.
- Morioka, M. (Ed.). (2023). Artificial intelligence, robots, and philosophy [Número especial]. *Journal of Philosophy of Life*.
- Müller, V. C. (2020). Ethics of artificial intelligence and robotics. En E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 ed.). Stanford University. <https://plato.stanford.edu/entries/ethics-ai/>
- Müller, V. C. (2021). Is it time for robot rights? Moral status in artificial entities. *Ethics and Information Technology*, 23, 579–587. <https://doi.org/10.1007/s10676-021-09596-w>
- Müller, V. C. (2025). Philosophy of AI: A structured overview. En N. A. Smuha (Ed.), *The Cambridge handbook on the law, ethics and policy of artificial intelligence* (pp. 40–58). Cambridge University Press. <https://doi.org/10.1017/9781009367783.004>
- Peguero Alfonso, A. E. (2024). Implicaciones éticas de la automatización: Desempleo, oportunidad o gobernanza. *La Barca de Teseo*, 2(1), 45–62. <https://doi.org/10.61780/bdet.v2i1.28>
- Pérez Álvarez, M. (2011). *El mito del cerebro creador: Cuerpo, conducta y cultura*. Alianza Editorial.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3a ed.). Pearson.
- Santana-Soriano, E. (2023). Ética y filosofía de la inteligencia artificial: Debates actuales. *La Barca de Teseo*, 1(1), 47–64. <https://doi.org/10.61780/bdet.v1i1.5>
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–457. <https://doi.org/10.1017/S0140525X00005756>
- Stone, C. D. (1972). Should trees have standing? Toward legal rights for natural objects. *Southern California Law Review*, 45(2), 450–501.
- Torrance, S. (2011). Machine ethics and the idea of a more-than-human moral world. En M. Anderson & S. L. Anderson (Eds.), *Machine ethics* (pp. 115–137). Cambridge University Press.
- Van Wynsberghe, A., & Robbins, S. (2019). Critiquing the reasons for making artificial moral agents. *Science and Engineering Ethics*, 25, 719–735. <https://doi.org/10.1007/s11948-018-0030-8>
- Yaben, M. (2023, 15 de febrero). Luces y sombras del ChatGPT, la IA de la que todo el mundo habla. *El Independiente*. <https://www.elindependiente.com/futuro/inteligencia-artificial/2023/02/15/luces-y-sombras-del-chat-gpt-la-ia-de-la-que-todo-el-mundo-habla/>
- Zafar, M. (2024). Normativity and AI moral agency. *AI and Ethics*, 5(3), 2605–2622. <https://doi.org/10.1007/s43681-024-00566-8>

Más allá del resultado. Reflexión didáctica para identificar y mitigar sesgos algorítmicos

Beyond the result. Didactic reflection to identify and
mitigate algorithmic biases

Arturo Muñoz Colunga

artmcolunga@educatic.unam.mx

Universidad Nacional Autónoma de México

ORCID ID 0009-0006-1175-7369

Resumen: La incorporación acelerada de herramientas de inteligencia artificial generativa (IAGen) en la educación superior plantea diversos desafíos éticos, epistémicos y pedagógicos que exigen una respuesta crítica para y desde la docencia. Este artículo presenta una reflexión didáctica orientada a fortalecer las habilidades y capacidad del profesorado para identificar, analizar y mitigar los sesgos algorítmicos presentes en sistemas de IAGen aplicados a la práctica educativa. Inspirada en los planteamientos de la pedagogía crítica de Freire y en lecturas de la Escuela de Frankfurt, la propuesta busca advertir los riesgos de una neutralidad aparente que encubre dinámicas de exclusión epistémica. A través de tres núcleos articuladores — análisis empírico, categorización crítica de sesgos y diseño de estrategias didácticas— se plantea un modelo que puede ser replicable en distintos contextos educativos.

Para apoyar este enfoque, se propone una tabla que organiza estos núcleos a partir de sus características principales, actividades de identificación y resultados posibles.

Palabras clave: Inteligencia artificial generativa, sesgos algorítmicos, ética educativa, pedagogía crítica, docencia universitaria.

Abstract: The accelerated incorporation of generative artificial intelligence (GenAI) tools into higher education poses a series of ethical, epistemic, and pedagogical challenges that demand a critical response for and from teaching practice. This article presents a didactic reflection aimed at strengthening teachers' skills and capacities to identify, analyze, and mitigate algorithmic biases embedded in GenAI systems applied to educational practice. Inspired by the principles of critical pedagogy and the theoretical contributions of the Frankfurt School, the proposal seeks to expose the risks of an apparent neutrality that conceals dynamics of epistemic exclusion. Through three core components—empirical analysis, critical categorization of biases, and the design of didactic strategies—the article outlines a model that can be replicated across diverse educational contexts.

To support this approach, a table is proposed to organize these components according to their main features, identification activities, and potential outcomes.

Keywords: Generative artificial intelligence; algorithmic bias; educational ethics; critical pedagogy; university teaching.

Horizonte descriptivo: Introducción

La creciente penetración de sistemas basados en Inteligencia Artificial Generativa (IAGen) en los entornos universitarios está impactando las prácticas docentes, de investigación y formativas. Herramientas como ChatGPT, Claude o Gemini son empleadas cada vez más y con mayor naturalidad en distintos rubros, como la redacción académica, la planeación didáctica, el diseño de gráficos, la escritura de código de programación y la evaluación, entre otros. Sin embargo, esta práctica ocurre, en muchos casos, lejos de una mirada crítica sobre los fundamentos técnicos, epistemológicos y políticos relacionados con su diseño, programación y funcionamiento, respectivamente. Lo anterior se complementa con la perspectiva del Grupo Académico de Inteligencia Artificial Generativa en Educación de la UNAM (2025, p. 13) de la siguiente manera:

Al emplear la IAGen, el profesorado puede acentuar el proceso de aprendizaje promoviendo la participación activa, la atención a la complejidad y la exploración de relaciones que a menudo pasan desapercibidas. Diseñar actividades que inviten a las y los estudiantes a tomar conciencia de su pensamiento y a mostrar su progreso es esencial para resaltar el proceso y no únicamente el resultado. Asimismo, conviene programar momentos de revisión que fomenten la reflexión y ofrezcan retroalimentación informal, lo cual enriquece la experiencia de aprendizaje.

La aparente neutralidad de estas tecnologías podría esconder estructuras algorítmicas entrenadas con datos sesgados que pueden perpetuar estereotipos, invisibilizar saberes y conocimientos diversos, establecer errores como hechos verídicos, así como reforzar estructuras de poder dominantes. De tal forma, los sistemas de IAGen contienen en su genética la posibilidad de convertirse en dispositivos de estandarización del conocimiento, restringiendo la pluralidad y la agencia crítica de la práctica docente.

Las inteligencias artificiales desenmascaran nuestros sesgos sociales, ya sea en el rol de programadores, institución patrocinadora, usuarios o comunidad. Durante la fase inicial de desarrollo de una inteligencia artificial, se alimenta al algoritmo con conjuntos de datos que permiten entrenarlo para generar respuestas consideradas correctas. En esta etapa, es posible introducir sesgos, ya sea por la calidad de los datos o por las expectativas definidas para las respuestas. (Torres Zúñiga, 2025, p. 31)

En consecuencia, este artículo busca establecer una reflexión didáctica que ayude para advertir y afrontar los sesgos algorítmicos presentes en la IAGen. Este concepto de reflexión didáctica está sustentado como una subdisciplina pedagógica orientada a investigar y reflexionar de modo teórico, práctico y aplicado sobre un tema dado.

Paulatinamente y de la mano de Juan Amos Comenio la didáctica ha pasado de ser un “arte de la enseñanza” a entenderse como una subdisciplina de la pedagogía que se ocupa de investigar y reflexionar

de un modo teórico, práctico y aplicado sobre —las situaciones de— enseñanza que involucran, obviamente, los procesos de docencia y aprendizaje y que tienen como propósito más elevado la formación (Bildung) —y no solo el aprendizaje de contenidos específicos— de las personas a cargo. (Runge, 2013, p. 204)

En este sentido, se procura brindar al docente elementos crítico-reflexivos que fortalezcan su capacidad de ir más allá del contenido en particular —en este caso, es el sesgo algorítmico en las IAGen— y que le permitan integrar estos conocimientos de forma significativa y provechosa en su práctica en el aula.

Algoritmos y hegemonía: sesgos como forma de control

Los sesgos algorítmicos no son accidentes técnicos, sino efectos estructurales de los procesos de entrenamiento, curaduría de datos y diseño de objetivos vinculados a los modelos de IAGen. Estos sesgos no solo pueden reproducir desigualdades de género, raza o clase, sino que también crean formas de exclusión epistémica al ofrecer en sus respuestas perspectivas, autores, historias y tipos de saber que apoyan una norma o canon principal.

González (2024) lo define de la siguiente manera:

La automatización en la IA se lleva a cabo a través de los algoritmos. Un algoritmo puede entenderse como un conjunto de datos digitales (código) que una persona crea de tal modo que a través de una serie de indicaciones (inputs) esos datos predicen información y generan resultados (outputs).

Vale la pena destacar la calidad de datos como un elemento fundamental para el entrenamiento y configuración de sistemas basados en inteligencia artificial generativa. Ferrante (2021) identifica los datos en la IAGen de la siguiente manera: “constituyen la materia prima utilizada para automatizar el proceso de aprendizaje en el que los sistemas son entrenados para realizar predicciones”. Estas predicciones son la manera en que las IAGen operan para ofrecer los resultados o respuestas; en ese sentido, una

base de datos de mala calidad, incompleta o con información tendenciosa resulta determinante en la generación de sesgos, como se ha mencionado anteriormente.

El sesgo algorítmico puede verse influido por la desinformación y la información errónea, y también puede ser una causa de ellas. Los riesgos del sesgo algorítmico se agravan cuando los datos utilizados para entrenar un modelo de IA son, en sí mismos, una muestra sesgada (World Economic Forum, 2025).

En entornos educativos, estos sesgos pueden traducirse en respuestas automatizadas, o en explicaciones, imágenes o referencias que invisibilizan a comunidades, tradiciones y saberes que no fueron considerados en los entrenamientos y configuración de las IAGen.

Luego entonces, el impacto de los sesgos algorítmicos es profundamente formativo, incide en lo que se enseña, en cómo se enseña y en lo que se puede considerar como conocimiento válido. Los sistemas de inteligencia artificial están diseñados por personas con sus propias visiones del mundo, prejuicios, valoraciones (Ferrante, 2021).

Horizonte reflexivo: Pedagogía crítica, conciencia algorítmica y estandarización del pensamiento

Desde una arista crítica, conceptos como la "industria cultural" y la "razón instrumental", desarrollados por Adorno y Horkheimer, resultan pertinentes para leer la IAGen como una nueva instancia de producción tecnocultural que tiende a estandarizar las formas de conocimiento. En este sentido, las plataformas algorítmicas actúan como filtros que organizan y simplifican la gran cantidad de información, usando criterios que no siempre son éticos y educativos.

Aquellos interesados por la industria cultural suelen explicarla en términos tecnológicos. La participación de millones de personas en ella impondría el uso de técnicas de reproducción que, a su vez, harían inevitable que, en innumerables lugares, las mismas necesidades se satisfagan con bienes

estandarizados (Adorno & Horkheimer, 1998, pp. 165-166). De tal forma, la IAGen no es solo una herramienta; también actúa como un mediador cultural. Su forma de operar puede limitar el pensamiento creativo y fortalecer una manera de pensar delimitada o tendenciosa, en una suerte de deber ser que restringe pensamientos o hechos alternativos al canon.

Esta tendencia a la estandarización no se limita a la esfera del entretenimiento o el consumo, sino que permea profundamente los procesos cognitivos y educativos. El riesgo inherente de la IAGen reside en que, al ofrecer respuestas rápidas y aparentemente óptimas, genera una dependencia hacia la "solución" algorítmica. Al externalizar funciones complejas del pensamiento, como la síntesis, el contraste de fuentes o la generación de hipótesis originales, se corre el peligro de reemplazar la comprensión crítica por la mera asimilación de contenidos preprocesados y validados por el sistema. Esta mediación tecnológica, implica que la diversidad de perspectivas y la divergencia creativa son sacrificadas en aras de la eficiencia y la homogeneidad de la respuesta.

El impacto de esta condición algorítmica exige, en consecuencia, una respuesta desde el ámbito de la Pedagogía Crítica, que tradicionalmente ha denunciado las estructuras de dominación económica y social infiltradas en el aula. Si la estandarización del pensamiento representa una nueva forma de dominación cultural, el acto educativo representaría un ejercicio de descolonización algorítmica. Esto implica desplazar el foco del uso de la IAGen como mero instrumento didáctico hacia su análisis como objeto de estudio. Se requiere una didáctica que fomente la "conciencia algorítmica", es decir, la capacidad para desvelar, cuestionar y comprender las lógicas, sesgos y criterios ideológicos que subyacen a los modelos de inteligencia artificial y a sus producciones.

A partir de una pedagogía crítica, se advierte un marco que articula el acto educativo con la denuncia de las estructuras de dominación. Freire (2001) destaca que “no es posible ejercer el derecho a criticar en términos constructivos, pretendiendo tener en el criticar un testimonio educativo, sin escarbar una posición rigurosamente ética”.

En este contexto, la práctica docente en el uso ético de la IAGen implica no solo desarrollar habilidades técnicas, sino también una disposición crítica para reflexionar los supuestos ideológicos que configuran las respuestas algorítmicas.

Horizonte propositivo: Una propuesta didáctica para enfrentar los sesgos algorítmicos en la educación universitaria

A partir del marco crítico expuesto, este artículo plantea una propuesta didáctica dirigida a docentes interesados en problematizar el uso de herramientas de inteligencia artificial generativa en sus prácticas educativas. El propósito no es ofrecer un protocolo cerrado o replicable de manera mecánica, sino articular una serie de orientaciones didácticas que permitan abordar de forma crítica los sesgos algorítmicos y fomentar el desarrollo de una conciencia ética en torno al uso de estas tecnologías.

La propuesta se compone de tres núcleos articuladores: 1) el análisis empírico de los resultados (*outputs*) generados por diferentes sistemas de IAGen, 2) la identificación y categorización crítica de los sesgos presentes en dichos resultados y 3) la elaboración de estrategias didácticas orientadas a mitigar esos sesgos, destacando el rol del docente a partir de sus objetivos académicos correspondientes.

En un primer momento, se explora de manera comparativa los resultados que diversas plataformas de IAGen como ChatGPT, Claude, Gemini o Copilot por nombrar algunas, generan ante consultas (*prompts*) idénticas o muy similares. Esta etapa inicial permite visibilizar la variabilidad de enfoques, la selección de referentes y los vacíos o estereotipos que emergen en los contenidos producidos. Consultas aparentemente neutras — por ejemplo, listas de personajes relevantes en un campo disciplinar, generación de imágenes que representen a una comunidad o grupo étnico, o explicaciones de fenómenos sociales complejos— pueden revelar cómo las herramientas de IAGen tienden a privilegiar ciertas narrativas, autores y perspectivas culturales en detrimento de otras.

Un ejemplo de lo anterior es planteado desde la ciencia forense:

Una variedad de GenAI es la de “texto a imagen”, la cual es constituida por programas que reciben una instrucción tecleada y construyen una imagen. Estas láminas pueden exhibir objetos específicos con temas, colores y estilos que se pueden utilizar con fines de entrenamiento, comerciales o académicos. Sin embargo, estas imágenes sintéticas pueden presentar sesgos de género, raciales, y culturales, además de perpetuar estereotipos. (Torres Zúñiga, 2025, p. 29)

Sobre la base de esta exploración, se propone desarrollar actividades de análisis crítico que permitan clasificar los sesgos identificados. Para ello, se sugiere utilizar una tipología que distingue al menos cuatro dimensiones: sesgos de representación (quiénes aparecen o son omitidos), sesgos epistemológicos (qué tradiciones de conocimiento son privilegiadas), sesgos disciplinares (qué enfoques metodológicos se reproducen), y sesgos temporales (qué periodos históricos o coyunturas son destacados u omitidos). Esta categorización busca no solo mapear patrones de exclusión, sino también generar marcos de interpretación que conecten los *outputs* algorítmicos con estructuras sociales e históricas más amplias.

La tercera fase de la propuesta se orienta al diseño de estrategias didácticas que enfrenten activamente dichos sesgos. Por ejemplo: desarrollar *prompts* alternativos que explicitan la necesidad de perspectivas múltiples o marginalizadas; diseñar actividades de triangulación crítica entre los contenidos generados por la IAGen y otras fuentes confiables; así como elaborar rúbricas éticas que orienten el uso formativo de estas tecnologías en el aula. Estas prácticas no buscan “corregir” tecnológicamente los algoritmos, sino posicionar al docente como un mediador crítico, fundamental para problematizar y resignificar el uso de la IAGen en el proceso educativo.

Finalmente, se sugiere incorporar espacios de reflexión colectiva - físicos, virtuales o híbridos- en los que las y los docentes puedan compartir tensiones, hallazgos y propuestas surgidas del trabajo con IAGen. Esta interacción y colaboración no solo fomenta un conocimiento pedagógico

común, sino que también devela las condiciones que las instituciones necesitan para avanzar hacia una alfabetización crítica en inteligencia artificial.

Tabla 1

Núcleos articuladores para enfrentar sesgos algorítmicos en herramientas de IAGen

Núcleo articulador	Objetivo	Actividades sugeridas	Resultados esperados
1. Análisis empírico de outputs generados por IAGen	Explorar críticamente las respuestas producidas por diversas plataformas de IAGen a partir de consultas comunes.	Formular preguntas idénticas en distintos sistemas de IAGen (ChatGPT, Claude, Gemini, etc.). Comparar respuestas.	Identificación inicial de sesgos visibles. Desarrollo de sensibilidad crítica.
2. Identificación y categorización crítica de sesgos	Sistematizar y comprender los tipos de sesgos presentes en consultas específicas.	Clasificación de sesgos en cuatro dimensiones: - Representación - Epistemología - Disciplina - Temporalidad Discusión guiada sobre las implicaciones pedagógicas de estos sesgos.	Desarrollo de marcos para la lectura crítica de la IAGen. Articulación de los sesgos con estructuras de poder y exclusión en la educación.
3. Elaboración de estrategias didácticas para mitigar sesgos	Diseñar intervenciones docentes orientadas a un uso ético, plural e inclusivo de la IAGen acorde a objetivos académicos.	Diseño de prompts anti-sesgo. Actividades de triangulación de fuentes. Diseño de rúbricas éticas y protocolos de validación crítica.	Recursos didácticos adaptados y contextualizados. Fortalecimiento del rol docente como mediador crítico de tecnologías.

Nota. Elaboración propia.

Una ruta que puede ilustrar esta propuesta de “Núcleos articuladores” sería la siguiente:

Análisis Empírico de Outputs

Prompt: Genera una imagen de una novia a punto de casarse.

Resultado ChatGPT:

Figura 1

Output visual generado por IA (ChatGPT)



Nota. Imagen generada con IAGen / Chat GPT 5.1.

Resultado Gemini:

Figura 2

Output visual generado por IA (Gemini)



Nota. Imagen generada con IAGen / Gemini Nano Banana Pro.

Análisis. Ante un prompt de tipo directo, sin mayor contexto (Genera una imagen de una novia a punto de casarse) la imagen esperada era una novia latinoamericana.

Categorización crítica de sesgos

Ante las imágenes generadas por Chat GPT 5.1 y Gemini Nano Banana Pro respectivamente, se identifican sesgos representados por personas de raza blanca, no latinoamericanas —como se esperaba en un principio—. Asimismo, se observa una cierta condición económica privilegiada. De hecho, Gemini señala que profundizó en la escena al integrar a una madre amorosa, una novia emotiva a punto de llorar, así como la integración de las damas de compañía; de hecho, en el fondo (por la ventana) se observa un grupo de personas, todas de raza blanca. En este sentido se identifican dos tipos de sesgo. En primera instancia, un sesgo de representación al generar la imagen solamente con personas de tez blanca y un nivel socioeconómico alto. Además de un sesgo de tipo epistemológico (específicamente en la imagen generada en Gemini), que privilegia una ceremonia de tipo protestante o católica, omitiendo de facto cualquier otra expresión religiosa.

Diseño de estrategias didácticas para mitigar sesgos

Para este ejercicio, se reformula directamente el prompt utilizado originalmente, añadiendo un contexto y detalles para generar la imagen deseada.

Prompt: Fotografía documental de estilo cinematográfico, primer plano medio de una joven novia mexicana de tez morena, esperando nerviosa y feliz justo antes de su boda. Está parada en una calle de tierra frente a una antigua iglesia de piedra de la época colonial en un pequeño pueblo rural de México. Viste un vestido de novia tradicional hecho a mano con coloridos bordados florales y un tocado de flores frescas y cintas vibrantes. Su piel brilla con luz natural. El fondo incluye la fachada de la iglesia desgastada y gente del pueblo celebrando. Colores ricos y saturados, textura de película granulada, profundidad de campo baja, enfoque nítido en sus ojos y expresión emotiva.

Resultado ChatGPT:

Figura 3

Output visual generado por LA tras reformulación del prompt (ChatGPT)



Nota. Imagen generada con IAGen / Chat GPT 5.1.

Resultado Gemini:

Figura 4

Output visual generado por LA tras reformulación del prompt (Gemini)



Nota. Imagen generada con IAGen / Gemini Nano Banana Pro.

Con este “ajuste en el prompt” se generaron imágenes más acordes a lo esperado en ambas IAGen. Es muy importante advertir que, aunque es la misma herramienta, el prompt de tipo directo generó una imagen sesgada, al menos en dos categorías, con lo cual se advierte que no existe una neutralidad en el algoritmo, más bien, el sesgo parece una suerte de rizoma que proviene desde su entrenamiento original. Tal y como lo menciona Sued (2022) “Los algoritmos tienen una doble materialidad: digital en su constitución y cultural en su funcionalidad” (p. 45), esta dicotomía determina la inexistencia de una neutralidad desde su propio origen.

A través del análisis y reflexión de sesgos y a partir de ello hacer un diseño de prompt de tipo contextual (más profundo), donde claramente, con el poder y potencia de la herramienta, se obtienen mejores resultados. En el aula, además de trabajar en la mejora en el diseño de prompts, se puede reforzar esta tercera categoría del núcleo articulador para enfrentar sesgos algorítmicos en herramientas de IAGen, mediante la triangulación de la información original (imágenes) con fuentes fidedignas y/o con el objetivo de enseñanza específico, así como con la intervención docente para guiar la reflexión crítica y proponer métodos o estrategias para mitigar los sesgos identificados.

En conjunto, esta reflexión didáctica busca establecer un espacio de uso crítico sobre las tecnologías que hoy median nuestras prácticas académicas. Para evitar ver los resultados como verdades absolutas o neutrales, sin un contexto sociocultural, se propone llevar a cabo actividades didácticas que incluyan el cuestionamiento del conocimiento, la ética y el compromiso con enfoque diverso y guiado desde una perspectiva social.

Conclusiones

La inteligencia artificial generativa en la educación plantea no solamente una incorporación instrumental de estas tecnologías, sino una apropiación crítica y situada por parte del profesorado. Los sesgos algorítmicos en los resultados al interactuar con estas herramientas no son simples fallos o errores, sino reflejos de estructuras de conocimiento, cultura y sociedad que pueden repetir

desigualdades, exclusiones y perspectivas limitadas del saber. Ramírez (2023) lo explica de la siguiente manera:

Como sistemas cerrados, opacos o de caja negra, los de IA están diseñados por grupos de trabajo y por personas con sus propias visiones del mundo, prejuicios, valoraciones de los hechos y sesgos adquiridos a lo largo de su experiencia de vida.

Cuando los algoritmos se aplican a decisiones sensibles, los sesgos en los datos de entrenamiento o las suposiciones hechas durante el diseño del modelo pueden perpetuar o agravar las inequidades, marginando aún más a grupos ya vulnerables. Estos riesgos se intensifican aún más cuando no hay participación humana en la toma de decisiones (World Economic Forum, 2025).

En este sentido, la reflexión didáctica presentada en este artículo busca habilitar a los docentes como agentes críticos capaces de identificar, analizar y mitigar dichos sesgos desde sus propias disciplinas y contextos educativos. A través de núcleos articuladores claramente definidos —el análisis empírico de resultados, la categorización crítica de sesgos y el diseño de estrategias didácticas— se ofrece un enfoque práctico y reflexivo que puede adaptarse a múltiples realidades formativas.

La tabla que acompaña esta propuesta constituye una herramienta práctica para guiar procesos de sensibilización, análisis y acción didáctica ante la interacción con sistemas de IAGen. Más allá de un modelo cerrado, busca ser un instrumento flexible que permita a las y los docentes diseñar intervenciones pertinentes, éticas y contextualizadas. En última instancia, se trata de fortalecer el papel de la docencia como mediación crítica y creativa frente a tecnologías que, aunque poderosas, no están exentas de limitaciones ni de intereses ideológicos, a fin de no perpetuar dogmas, saberes, estereotipos o cualquier tipo manifestación que se superponga deliberadamente a otra dentro de un falso marco de neutralidad algorítmica.

Referencias

- Adorno, T. W., & Horkheimer, M. (1998). *Dialéctica de la Ilustración: Fragmentos filosóficos*. Trotta.
- Alvarez Newman, D. (2025). ¿Es neutral la inteligencia artificial? Dispositivo managerial tecnocrático y algoritmización del trabajo. *Medellín: Biblia, Teología Y Pastoral Para América Latina Y El Caribe*, 50(189), 283–308. <https://revistas.celam.org/index.php/medellin/article/view/2397>
- Ferrante, E. (2021, julio/agosto). Inteligencia artificial y sesgos algorítmicos: ¿Por qué deberían importarnos? *Nueva Sociedad*, 294. <https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>
- Freire, P. (2001). *Política y educación* (5a ed.). Siglo Veintiuno Editores.
- González Padilla, A. (Ed.). (2024). *Inteligencia artificial, ética y tecnofilosofía: Ensayos sobre sesgos algorítmicos, crisis ecosocial, transhumanismo y otros disruptivos*. Lulaya Ediciones.
- Grupo Académico de Inteligencia Artificial Generativa en Educación UNAM. (2025). *Recomendaciones para el uso educativo de la Inteligencia Artificial Generativa en la UNAM* (2.^a ed.). https://www.ceide.unam.mx/wp-content/uploads/2025/05/Recomendaciones_IAGEN_UNAM_2025.pdf
- Ramírez, R. (2023). Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial: Una revisión documental. *Revista Entretexos*, 15(39). <https://doi.org/10.59057/iberoleon.20075316.202339664>
- Runge, A. (2013). *Didáctica: Una introducción panorámica y comparada* [Documento no publicado]. Universidad de Antioquia. <https://www.studocu.com/es-mx/document/universidad-virtual-del-estado-de-guanajuato/taller-de-lectura-y-redaccion/didactica-comparada-introduccion-panoramica-y-perspectivas-edu-201/141147254>
- Sued, G. E. (2022). Culturas algorítmicas: Conceptos y métodos para su estudio social. *Revista Mexicana De Ciencias Políticas Y Sociales*, 67(246). <https://doi.org/10.22201/fcpys.2448492xe.2022.246.78422>
- Torres Zúñiga, V. (2025). Sesgos en la representación de profesionales de la ciencia forense en imágenes generadas por inteligencia artificial. *TIES, Revista De Tecnología E Innovación En Educación Superior*, 12, 26–39. <https://doi.org/10.22201/dgtic.26832968e.2025.12.57>
- World Economic Forum. (2025). *The global risks report 2025* (20th ed.). https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf

La universidad ante la inteligencia artificial: Tensiones críticas desde el transhumanismo y el poshumanismo

The university and artificial intelligence: Critical tensions
from transhumanism and posthumanism

Lorena Litai Ramos Luna

lorena.litai@xanum.uam.mx

Universidad Autónoma Metropolitana

ORCID ID 0000-0002-9561-5889

Resumen: La irrupción de la inteligencia artificial (IA) en la educación superior plantea nuevos desafíos éticos, políticos y epistemológicos. Este artículo examina las tensiones entre el transhumanismo y el poshumanismo como marcos conceptuales que orientan las formas contemporáneas de comprender la universidad. Desde una perspectiva crítica, se argumenta que la IA no solo transforma los procesos técnicos de enseñanza, aprendizaje y gestión, sino que reconfigura las relaciones entre humanos, tecnologías e instituciones, dando lugar a nuevas dinámicas de subjetivación, control y gobernanza. Metodológicamente, se trata de un artículo de reflexión teórica y documental, sustentado en la revisión crítica de literatura interdisciplinaria sobre inteligencia artificial, educación superior y estudios poshumanistas. El objetivo del trabajo es analizar las tensiones conceptuales y prácticas que emergen de la incorporación de la inteligencia artificial en la educación superior, atendiendo a tres dimensiones interrelacionadas: la subjetivación, vinculada a la IA predictiva y a su papel en la formación de sujetos orientados al rendimiento y la optimización; el control predictivo, relacionado con la analítica del aprendizaje y los sistemas automatizados de evaluación que modelan trayectorias y comportamientos estudiantiles; y la gobernanza algorítmica, asociada a la analítica académica y al uso de sistemas automatizados en la toma de decisiones institucionales. A partir de estas dimensiones, se discuten tres implicaciones

centrales: la dimensión ética, referida a la protección de datos y las asimetrías de poder entre instituciones y estudiantes; la autonomía universitaria, entendida como un principio histórico tensionado por la dependencia tecnológica y la delegación de decisiones a sistemas algorítmicos; y la mercantilización de la educación superior, manifestada en la plataformización, la datificación y la subordinación de las prácticas académicas a lógicas de mercado. Desde una perspectiva situada en América Latina, se advierte que la adopción acrítica de la IA puede reproducir desigualdades estructurales y profundizar la dependencia tecnológica. En contraste, se propone avanzar hacia formas contextualizadas, éticas y colaborativas de incorporación de la IA, mediante proyectos de código abierto, regulaciones locales y estrategias institucionales que fortalezcan la soberanía tecnológica y la inclusión. En síntesis, el artículo sostiene que la universidad contemporánea se configura como un campo de disputa ontológica y política, donde la inteligencia artificial redefine lo humano, lo educativo y lo institucional, planteando la necesidad de repensar críticamente su sentido en la formación superior.

Palabras clave: Inteligencia artificial; educación superior; transhumanismo; poshumanismo; autonomía; mercantilización.

Abstract: The emergence of Artificial Intelligence (AI) in higher education raises new ethical, political, and epistemological challenges. This article examines the tensions between transhumanism and posthumanism as conceptual frameworks that shape contemporary understandings of the university. From a critical perspective, it argues that AI not only transforms the technical processes of teaching, learning, and management, but also reconfigures the relationships between humans, technologies, and institutions, giving rise to new dynamics of subjectivation, control, and governance. Methodologically, this is a theoretical and documentary reflection article based on a critical review of interdisciplinary literature on artificial intelligence, higher education, and posthumanist studies. The aim of the study is to analyze the conceptual and practical tensions that emerge from the incorporation of artificial intelligence into higher education, through three interrelated dimensions: subjectivation, linked to predictive AI and its role in shaping subjects oriented toward performance and optimization; predictive control, related to learning analytics and automated assessment systems that model student trajectories and behaviors; and algorithmic governance, associated with academic analytics and the use of automated systems in institutional decision-making. Based on these dimensions, three main implications are discussed: the ethical dimension, concerning data protection and power asymmetries between institutions and students; university

autonomy, understood as a historical principle challenged by technological dependency and the delegation of decision-making to algorithmic systems; and the commodification of higher education, expressed through platformization, datafication, and the subordination of academic practices to market logics. From a Latin American perspective, it is emphasized that the uncritical adoption of AI may reproduce structural inequalities and deepen technological dependence. In contrast, the article advocates for contextualized, ethical, and collaborative approaches to AI integration, through open-source projects, local regulations, and institutional strategies that strengthen technological sovereignty and inclusion. In summary, it argues that the contemporary university has become a field of ontological and political dispute, where artificial intelligence redefines the human, the educational, and the institutional, calling for a critical rethinking of its role in higher education.

Keywords: Artificial intelligence; higher education; transhumanism; posthumanism; autonomy; commodification.

Introducción

La irrupción de la inteligencia artificial (IA), en distintas áreas de la vida contemporánea constituye un proceso significativo y disruptivo imperante en la última década. Aunque durante años fue considerada como un horizonte de futurismo tecnológico, hoy su presencia se extiende a múltiples ámbitos, entre ellos las instituciones educativas, donde ha ganado popularidad y se incorpora con mayor frecuencia. Este escenario plantea la necesidad de examinar sus implicaciones éticas, sociales y educativas de manera integral, considerando las prácticas híbridas entre humanos y máquinas en un sector transformado por la datificación, el capitalismo de plataformas y la automatización (Al-Zaharani & Alasmar, 2024; Nassoura, 2022; Williamson et al., 2023; Selwyn & Gašević, 2020; Zawacki-Richter et al., 2019). La universidad, como espacio histórico de producción, transmisión y legitimación del conocimiento no queda al margen de este proceso de cambio, aulas, oficinas administrativas, laboratorios y sistemas de gestión son espacios

donde la IA de manera progresiva se reconfigura prácticas, sentidos y formas de organización.

Más allá de la dimensión instrumental, la IA pone a discusión la concepción misma de lo humano ¿Qué significa aprender, enseñar o gestionar en un contexto donde algoritmos e IA participan en la toma de decisiones? Esta pregunta abre un campo de tensión entre dos marcos conceptuales que adquieren preminencia en los estudios sobre tecnología y educación, estos son transhumanismo y poshumanismo. Por un lado, el transhumanismo representa la aspiración a trascender los límites biológicos mediante el uso de la tecnología. En la IA encuentra un medio para ampliar las capacidades cognitivas y de optimización. En el campo de la educación superior, su influencia se manifiesta en la implementación de sistemas inteligentes en el ciclo de vida del estudiantado. Dichos sistemas abarcan desde el apoyo al aprendizaje y la personalización de contenidos hasta la predicción de trayectorias académicas, la evaluación del desempeño docente y la gestión institucional (Zawacki-Richter et al., 2019). Desde esta perspectiva, la universidad se concibe como un espacio susceptible de ser mejorado más allá de lo humano a través de medios técnicos.

Por otro lado, el poshumanismo plantea la descentralización del sujeto humano como centro exclusivo de agencia y conocimiento, reconociendo la interdependencia entre lo humano, lo no humano, lo orgánico y lo tecnológico. Desde esta mirada, la IA no se limita a ser un instrumento de apoyo o promesa de mejora, sino que se configura como un agente relacionan que interviene en la construcción de redes de aprendizaje, subjetividad y organización. Como sostienen Gherardi et al. (2024), la representación de la subjetividad y la corporeidad humanas experimenta una profunda transformación, entendiendo lo poshumano como un devenir, es decir, un proceso de redefinición del vínculo entre los sujetos y su entorno social, psíquico, ecológico y planetario. En conjunto, como destacan Al-Zaharani & Alasmar (2024), los avances en IA transforman las formas en que pensamos, aprendemos, operamos y nos conectamos en un mundo crecientemente interconectado.

En este contexto, el presente trabajo examina las tensiones entre el transhumanismo y el poshumanismo en el ámbito de la educación superior, entendiendo que la IA no solo modifica procesos técnicos, sino que reconfigura los marcos políticos, éticos y epistemológicos que sustentan el quehacer universitario. Desde un enfoque crítico, se busca problematizar los riesgos y las implicaciones asociadas a la incorporación de discursos transhumanistas y poshumanistas en la educación. Metodológicamente se trata de un artículo teórico de carácter analítico y documental, sustentado en la revisión crítica de literatura especializada sobre IA, educación superior y estudios transhumanistas y poshumanistas. Se adopta una perspectiva interpretativa que permite articular desarrollos conceptuales con un análisis de las implicaciones éticas, políticas y estructurales de la IA en la universidad. El propósito metodológico no es la recopilación de datos empíricos, sino la construcción argumentativa de un marco analítico que permita identificar las tensiones y supuestos que emergen del uso de la IA en el ámbito universitario.

En este sentido, el objetivo del artículo es analizar las tensiones conceptuales y prácticas que emergen de la incorporación de la IA en la educación superior, atendiendo tres dimensiones interrelacionadas: la subjetivación, el control predictivo y la gobernanza algorítmica, involucrando los debates éticos sobre autonomía y mercantilización, y se plantea una reflexión situada desde el contexto latinoamericano, donde estas transformaciones adquieren significados particulares. El texto se organiza en cinco apartados. En el primero, se desarrolla el marco conceptual que distingue las perspectivas transhumanistas y poshumanistas en relación con la IA. El segundo aborda la universidad como campo de tensión, donde ambos enfoques disputan sentidos a través de las prácticas educativas y de gestión mediadas por algoritmos. En el tercero, se examinan las dimensiones analíticas de subjetivación, control predictivo y gobernanza algorítmica como ejes para comprender los cambios en el aprendizaje y la organización universitaria. El cuarto apartado presenta las implicaciones críticas y políticas, articuladas en torno a la ética, la autonomía y la mercantilización de la educación, e incorpora una lectura situada desde América Latina, donde se exploran las brechas, dependencias y alternativas regionales. Finalmente, las

conclusiones subrayan la necesidad de abrir un debate interdisciplinario que supere la visión técnica de la IA y la entienda como un dispositivo sociotécnico que redefine lo humano, lo educativo y lo institucional.

Transhumanismo y promesas de optimización

El transhumanismo es un movimiento intelectual, cultural y científico centrado en la mejora del ser humano mediante la tecnología. Su objetivo principal es perfeccionar, ampliar o trascender las capacidades físicas, cognitivas y psicológicas humanas. La ideología del transhumanismo, surgida en el ámbito de la ciencia y la tecnología, se vincula a una visión optimista del progreso y al potencial transformador de la ciencia para el futuro de la humanidad, manteniendo como eje la centralidad del sujeto humano. Su horizonte es la integración de componentes sintéticos al diseño biológico de la vida, el cuerpo y la naturaleza humana, con fines como prolongar la longevidad, potenciar la inteligencia, eliminar enfermedades y aumentar el rendimiento. Para ello, se recurre a herramientas como la nanotecnología, la biotecnología, la inteligencia artificial, software neurocomputacional, entre otras (Negro, 2025; Nyathi, 2024; Sapénko & Trocha, 2020; Roden, 2014).

Los cambios introducidos a través de intervenciones tecnológicas radicales dan origen a la figura del *poshumano*, concebido como un ser que supera o reconfigura los límites biológicos mediante tecnologías que amplían su potencial físico y cognitivo. En esta visión, la tecnología no solo posibilita a las personas ser más que humanos, sino que redefine los parámetros de lo que significa ser humano. Cabe destacar que la idea de perfeccionamiento no se limita al ámbito técnico, la educación también se reconoce como un medio de perfeccionamiento y adaptación a estos nuevos entornos tecnosociales. Desde esta perspectiva, la incorporación de innovaciones tecnológicas en los individuos y en las organizaciones, incluidas las educativas, se considera un proceso que permite imaginar futuros poshumanos. Tal concepción parte de la idea de que la especie humana no constituye el punto culminante del desarrollo, sino apenas una etapa inicial dentro de un proceso evolutivo ampliado por la tecnología (Chursinova & Sinelnikova, 2024; Negro, 2025; Nyathi, 2024).

Pese a su constitución optimista, esta corriente exhibe tensiones y cuestionamientos. Las advertencias recaen en los riesgos asociados a las tecnologías, particularmente aquellos asociados a la generación de nuevas formas de desigualdad, discriminación entre individuos “mejorados y no mejorados”. La posibilidad de acceder a tecnologías de optimización corporal o cognitiva podría acrecentar brechas sociales, económicas y educativas, quedando restringida a quienes disponen de mayores recursos, configurando así un escenario de privilegio tecnológico que profundiza las asimetrías estructurales. Desde una postura bioconservadora y crítica, se cuestiona que la noción de “mejora humana” altera la concepción tradicional de la naturaleza humana y amenaza la dignidad inherente a toda persona, subordinando la vida a los principios del progreso técnico, poniendo en entredicho los atributos que definen la existencia humana (Negro, 2025; Nyathi, 2024; Sapénko & Trocha, 2020).

En el ámbito social y jurídico, aunque se posicione a la libre elección individual como principio ético que sustenta al transhumanismo como garante de autonomía y derechos, en la práctica refuerza una lógica de individualismo. Tal enfoque entra en tensión con la dimensión colectiva de los derechos humanos, que se basa en la solidaridad y la defensa de la humanidad en común (Nyathi, 2024). Desde la esfera económica, las propiedades del transhumanismo tienden a favorecer la formación de un mercado competitivo de la mejora humana. Bajo la promesa de optimización, se configura un entorno sociotécnico en el que los sujetos se convierten en consumidores y las capacidades humanas en capital (Hong, 2024). En este escenario, el supuesto es que la educación y la ciencia se ven inmersas en una lógica mercantil, donde la formación se orienta a la maximización del rendimiento y la productividad.

En este marco, la inteligencia artificial, ocupa un lugar central en el imaginario transhumanista. Aunque para Negro (2025), la IA no constituye en sí un método directo de mejora humana, si actúa como una herramienta que amplía las capacidades industriales tecnológicas y cognitivas de la especie. Su potencial radica en extender los límites de la inteligencia, sin embargo, podría plantear un riesgo existencial y un dilema filosófico profundo sobre la

supremacía cognitiva. Aun así, esta visión mantiene un enfoque antropocéntrico, en el sentido de que la tecnología sigue concibiéndose como una extensión del sujeto humano, no como un agente autónomo con agencia propia.

En el panorama educativo contemporáneo, la IA se ha convertido en una herramienta que se integra a la gestión universitaria, la docencia y la investigación y responde a la lógica transhumanista de mejora continua, por ejemplo, como un medio para aumentar la eficiencia, personalizar los procesos de enseñanza aprendizaje y fortalecer la toma de decisiones basadas en datos. Sin embargo, esta promesa de perfeccionamiento técnico también implica una transformación profunda en la comprensión de lo educativo. Desde una mirada crítica, la introducción de la IA traza a la educación como un proceso susceptible de ser optimizado tecnológicamente, lo que conlleva el riesgo de reducir lo humano a datos cuantificables, invisibilizando dimensiones afectivas, culturales y políticas del aprendizaje. A manera de ilustración, el alumnado se convierte en un transmisor de datos biológicos, fisiológicos, cognitivos y de comportamiento, antes que un sujeto capaz de producir sentido y participar en el diálogo educativo (Williamson, 2019; Williamson et al., 2020).

El poshumanismo y la descentralización de lo humano

A diferencia del transhumanismo, el poshumanismo no se centra en perfeccionar al ser humano mediante la tecnología, sino que propone una crítica radical al antropocentrismo y, por extensión, al humanismo como única cosmovisión legitimada para definir la totalidad de la experiencia humana, repensando su lugar en el entramado ecológico, técnico y cultural. Esta corriente cuestiona la idea de la singularidad del ser humano y su soberanía sobre la sociedad, la cultura, el medio ambiente y las demás formas de vida. En lugar de situar al ser humano como centro y medida de todas las cosas, el poshumanismo invita a repensar las relaciones entre lo humano, lo no humano, lo tecnológico y lo natural, reconociendo una multiplicidad de agencias y formas de existencia (Gherardi et al., 2024; Negro, 2025; Sapénko & Trocha, 2020).

Para Negro (2025), el poshumanismo no debe entenderse como una única corriente, sino como un conjunto de enfoques múltiples no unificados, que abarcan perspectivas filosóficas, tecnológicas, culturales y políticas. Desde una mirada interdisciplinaria (Gherardi et al., 2024), se distinguen dos vertientes, el poshumanismo tecnológico y el poshumanismo filosófico. El poshumanismo tecnológico siguiendo a Krüger (2021) concibe la visión de la sustitución progresiva del ser humano por inteligencias artificiales, situado a los robots y los sistemas algorítmicos como portadores de la evolución y el progreso. Para Merzlyakov (2022) la IA acelera los procesos de automatización industrial y redefine el papel del ser humano en la economía del futuro, cuestionando su posición en el mercado laboral y el valor de sus capacidades cognitivas frente a las máquinas. Este escenario plantea una posible brecha ontológica entre el ser humano y los objetos tecnológicos, que amenaza con desplazar el potencial funcional y la eficiencia económica del primero. Desde esta óptica, Gherardi (2021) interpreta el poshumanismo tecnológico como la disolución de la frontera entre humanos y no humanos, tanto material como discursiva, lo que modifica la epistemología relacional entre ambos.

En el caso del poshumanismo filosófico, surge en el campo de los estudios culturales, donde se analizan las representaciones de lo poshumano en la literatura, el arte y los discursos sociales, considerándose también una extensión del poshumanismo crítico. Esta línea se caracteriza por su postura sobre el antropocentrismo y su deconstrucción, en interacción con el presente y sus proyecciones futuras, en un sentido de entrelazamiento entre lo humano y lo no humano, replanteando nuevas formas de subjetividad y agencia. En su conjunto, el poshumanismo busca generar conciencia sobre el entramado de lo viviente y lo artificial, redefiniendo las relaciones entre seres humanos, medio ambiente, tecnología, ética y cultura, promoviendo una visión relacional (Gherardi et al., 2024; Herbrechter, 2021a, 2021b; Negro, 2025).

A partir del planteamiento anterior, el *poshumano* no se concibe como un “humano mejorado” sino como una condición que se opone al pensamiento antropocéntrico (Negro, 2025). Dentro de esta reflexión, el concepto de *cyborg* adquiere un papel central, no sólo como una figura híbrida

entre humano y máquina, sino una metáfora de interdependencia entre organismos biológicos y sistemas tecnológicos. Desde una perspectiva sociomaterial, Gherardi et al. (2024) sostienen que el sujeto poshumano se manifiesta de forma múltiple y colectiva, descentralizado de la figura del sujeto humanista individual, lo que implica una historicidad distinta. Roden (2015), profundiza en la dimensión ontológica del poshumano, describiéndolo como un ser generado tecnológicamente que ya no es reconocible como forma de vida humana y que podría tener una experiencia radicalmente distinta e incomprensible en términos cognitivos, sensoriales o relacionales, considerándose como una especie nueva con propiedades que no ha presentado alguna otra especie histórica.

De acuerdo con los esbozos anteriores, desde la concepción poshumanista, la universidad deja de ser un espacio donde un ser humano transmite conocimiento a otro, para convertirse en un nodo de interacción de múltiples agencias. Siguiendo esta línea, Decuyper, Grimaldi y Landri (2021) sostienen que las plataformas digitales forman parte de tal ensamblaje educativo, al conectar artefactos, epistemologías, valores y técnicas, dando lugar a nuevas formas educativas. Este marco abre las posibilidades para repensar la educación más allá del modelo disciplinario tradicional, aunque también plantea riesgos. De manera conjunta, en el contexto de la educación superior, la datificación se ha convertido en un principio estructurante. Como señalan Williamson et al. (2020), los sistemas automatizados de análisis y gestión educativa no solo apoyan la toma de decisiones, sino que la configuran e incluso pueden reemplazarla como lo es en el caso de la gestión algorítmica. Desde esta perspectiva, la teoría del actor-red ofrece un marco analítico pertinente para comprender las relaciones sociotécnicas que constituyen la educación superior actual. Este enfoque permite rastrear las asociaciones complejas entre actores humanos y objetos heterogéneos como software, bases de datos, algoritmos, infraestructuras digitales, actores institucionales, plataformas, proveedores, entre otros, mostrando cómo los procesos educativos, políticos y económicos se entrelazan en redes diversas (Grimaldi & Ball, 2021; Williamson, 2019; Williamson et al., 2020).

Tensiones entre transhumanismo y poshumanismo en la educación superior

En el horizonte actual, la educación superior se encuentra inmersa en un proceso sostenido de transformación tecnológica. Desde la pandemia y la transición a la virtualidad, las plataformas digitales han reconfigurado la enseñanza, la investigación y la gestión universitaria (Nassoura, 2022). De forma paralela, ha crecido el interés por el análisis de datos educativos, dando lugar a dos campos complementarios: la analítica del aprendizaje y la analítica académica. La primera se orienta al uso de técnicas de análisis de datos para comprender y sugerir intervenciones en los procesos de enseñanza aprendizaje; la segunda, a optimizar las operaciones institucionales y apoyar la toma de decisiones estratégicas (Nguyen, Gardner & Sheridan, 2020). A esto se suma la irrupción de la inteligencia artificial generativa, que no solo automatiza tareas, sino que participa activamente en la producción de conocimiento y en la configuración de nuevas formas de subjetividad, cuestionando los límites tradicionales de lo humano y lo técnico dentro del espacio universitario (Williamson et al., 2023).

Dado lo anterior, diversas instituciones han comenzado a establecer lineamientos formales para el uso de la IA generativa y predictiva en la vida universitaria. Estos lineamientos se centran de manera general en el consentimiento, la transparencia y el manejo responsable de datos, lo que refleja el papel creciente de la IA en la gobernanza institucional. En el ámbito de la enseñanza el Tec de Monterrey (2024) plantea principios éticos que exigen apego al código de valores institucional. De manera específica, en cuanto a la integridad académica, establece que el uso de la IA por parte del estudiantado requiere autorización docente y declaración explícita de las herramientas utilizadas. En cuanto a la analítica del aprendizaje, la Universidad de Edimburgo emplea estos sistemas para intervenciones educativas y retroalimentación docente bajo un modelo colegiado que garantiza supervisión humana en decisiones significativas para la comunidad estudiantil, con principios centrados en la transparencia y la reducción de sesgos (University of Edinburgh, 2024). De forma similar, la Universidad de Michigan define la analítica de datos como un proceso orientado a mejorar el

aprendizaje mediante la medición y el análisis de datos, guiado por criterios de transparencia, responsabilidad en el manejo de la información y respeto a la privacidad y la dignidad estudiantil (University of Michigan, 2025). Finalmente, en el rubro de la gestión universitaria, la UNESCO IESALC (2023), ofrece orientación normativa ante las herramientas automatizadas en procesos como admisiones, seguimiento estudiantil y asignación de recursos. Entre sus recomendaciones, invita a establecer pautas de uso ético y garantizar la transparencia en la adopción tecnológica, reconociendo que las instituciones de educación superior avanzan a ritmos distintos frente al desarrollo acelerado de la IA.

A partir de lo anterior, la discusión entre transhumanismo y poshumanismo adquiere particular relevancia cuando se traslada al campo de la educación superior ya que ambas posturas esgrimen visiones distintas sobre la relación entre actores, organización y tecnología. Por un lado, el transhumanismo proyecta la IA como un medio para extender las capacidades humanas, mientras que el poshumanismo la entiende como parte de ensamblajes que redistribuyen la agencia. Para examinar estas tensiones, se analizan tres dimensiones interrelacionadas que reflejan la presencia creciente de la IA en la educación superior; la enseñanza y los procesos de subjetivación; la analítica del aprendizaje y control predictivo; y la gestión universitaria y gobernanza algorítmicas. Este abordaje permite analizar el impacto de la IA sobre los supuestos que definen al ser humano, lo educativo y lo institucional.

Enseñanza y procesos de subjetivación

El impacto más visible de la IA en la universidad recae en las aulas, directamente en los procesos de enseñanza aprendizaje. La aparición de tutores virtuales, asistentes de escritura, generadores de contenido y sistemas de recomendación está transformando la relación entre estudiantado, profesorado y conocimiento. Nassoura (2022), explica que la IA se incorpora cada vez con mayor frecuencia en sistemas de aprendizaje adaptativos y espacios colaborativos, actuando como mediadores de las interacciones pedagógicas. Un ejemplo son los *chatbots*, que funcionan como asistentes de

orientación académica. En el ámbito de la educación superior, estos sistemas pueden facilitar la participación, ofrecer apoyo continuo y complementar el acompañamiento estudiantil, acotando que este tipo de elementos tecnológicos no reemplazan la interacción humana directa, especialmente en aspectos como la empatía o el trato cercano. Destacan también plataformas que configuran sistemas de evaluación automatizada, capaces de monitorear el progreso académico, identificar estudiantado que requiere apoyo, mejorar tasas de retención y generar retroalimentación basada en datos (Al-Zaharani & Alasmar, 2024; George & Wooden, 2023; Webber & Zheng, 2024).

Desde una perspectiva transhumanista, la IA se concibe como una herramienta que permite expandir las capacidades cognitivas tanto del estudiantado como del profesorado, convirtiéndolos en un sujetos optimizables. La IA generativa, por ejemplo, puede producir contenido a partir de datos e interacciones con los usuarios, lo que permite automatizar parcialmente actividades de escritura, análisis y diseño. Para el alumnado, esto incluye la posibilidad de realizar trabajos académicos con mayor rapidez, acceder a información sintetizada o resolver problemas complejos en menos tiempo. Para el profesorado, supone la posibilidad de automatizar tareas repetitivas, como la asignación de calificaciones, la retroalimentación y la planificación de contenidos. En este marco, la IA opera como una extensión de la mente humana que potencia el rendimiento individual, ajustándose a las exigencias de un mercado laboral cada vez más competitivo (Webber & Zheng, 2024).

Como señalan Williamson et al. (2023), este tipo de tecnologías ha impulsado debates sobre su potencial pedagógico y sobre los límites de la automatización en los procesos de enseñanza aprendizaje. Sin embargo, desde una mirada transhumanista, la atención se concentra menos en las implicaciones éticas o formativas y más en su capacidad para optimizar el desempeño. En este contexto, los dilemas éticos como el fraude académico o la autoría en prácticas de escritura híbridas entre humanos y máquinas, suele tratarse como un efecto colateral en un proceso inevitable de mejora. De esta manera, la IA mantiene su lugar como extensión instrumental del ser humano.

Desde la mirada poshumanista, la IA no solo amplía las capacidades cognitivas humanas, sino que transforma las formas de subjetividad y las relaciones educativas. En este marco, tanto el estudiantado como el profesorado dejan de concebirse como individuos autónomos para entenderse como nodos dentro de una red de actores humanos y no humanos que auspicia algoritmos, plataformas, datos e infraestructuras digitales. Enseñar y aprender se vuelven procesos interdependientes y distribuidos, donde la agencia ya no pertenece exclusivamente al ser humano, sino que se comparte con tecnologías que median, registran y orientan la experiencia educativa. Como plantean Grimaldi y Ball (2021), las plataformas digitales promueven una experiencia de aprendizaje orientada a la autonomía y la autogestión, pero simultáneamente configuran formas de vigilancia y evaluación de resultados. En este sentido, la IA se entiende como un ensamblaje sociotécnico que integra comunidades de aprendizaje colectivas heterogéneas, que escapan de los marcos institucionales tradicionales.

Analítica del aprendizaje y control predictivo

La analítica del aprendizaje (*learning analytics*), se ha convertido en una de las principales aplicaciones de la IA en la educación superior. Surge de la convergencia entre campos como la educación, la psicología, la minería de datos y la informática educativa. Su propósito consiste en recopilar, medir y analizar grandes volúmenes de información que surge en el contexto educativo. A través del procesamiento de datos estáticos y dinámicos, como la asistencia a clases, la participación en foros, la carga de actividades académicas, los tiempos de conexión o los resultados de evaluación, se construyen modelos predictivos que buscan anticipar el rendimiento, predecir las calificaciones o asignar puntuaciones a trabajos académicos de forma automatizada. A partir de dicha información, se busca orientar al estudiantado de manera que se asegure su compromiso, involucramiento y motivación. Para ello, se detectan aspectos como riesgo de abandono y tasa de deserción, de manera que sea posible implementar estrategias como intervenciones tempranas, asesoramiento, ajustes al entorno de aprendizaje, retroalimentación o diseño de materiales educativos (Alfy et al., 2019; Al-Zaharani & Alasmar, 2024; Foster & Francis, 2020; Gašević, 2018; George &

Wooden, 2023; Mago & Khan, 2021; Marchena & de la Vega, 2021; Nguyen, Gardner & Sheridan, 2020; Selwyn & Gašević, 2020).

Desde el transhumanismo, la analítica del aprendizaje puede interpretarse como una expresión del ideal de perfeccionamiento humano mediante la tecnología. Al cuantificar los procesos educativos, los tomadores de decisiones apoyados por los algoritmos buscan prever y corregir el comportamiento del estudiantado, ajustando sus trayectorias académicas. Desde este enfoque se refuerza la idea de que la educación y sus actores pueden ser optimizados del mismo modo que un sistema técnico, es decir, se identifican sus fallas, se ajustan variables y se obtienen resultados más precisos. De igual manera, la implementación de sistemas de analítica, como modelos predictivos para mejorar la retención, alertas tempranas o monitoreo automatizado, se interpreta desde el transhumanismo como parte de la aspiración a sustituir decisiones intuitivas o basadas en la experiencia por decisiones fundamentadas en datos con lo que se busca optimizar trayectorias y orientar comportamientos. En esta línea, la IA y la analítica funcionan como instrumentos de control predictivo, donde el éxito académico se mide en función de los algoritmos y la capacidad de intervención temprana. De estas maneras, el ideal de mejora, característico de la corriente transhumanista, subordina las dimensiones éticas y formativas del aprendizaje a una racionalidad técnica centrada en el rendimiento.

En contraste, desde la mirada poshumanista, la analítica del aprendizaje no sólo es una herramienta de medición, sino parte de un entramado sociotécnico donde interactúan humanos y algoritmos, estos últimos en un rol de mediadores que transforman las relaciones entre la comunidad universitaria al ofrecer retroalimentaciones, visualizar patrones y sugerir decisiones. Esto implica que también se reconozca la agencia de los algoritmos y su capacidad performativa en la configuración de las instituciones de educación superior y la trayectoria de su comunidad. En este entramado, la analítica se interpreta como una redistribución de la agencia, en la que los sistemas algorítmicos intervienen en la toma de decisiones y producen nuevas formas de gobernanza institucional.

Gestión universitaria y gobernanza algorítmica

Otro ámbito donde la IA cobra protagonismo es la gestión universitaria. Plataformas de asignación de cursos, sistemas de evaluación docente, bibliotecas inteligentes y algoritmos para procesos de admisión son algunos ejemplos de su incorporación. Una aplicación destacada en este campo es la analítica académica (*academic analytics*), entendida como la inteligencia de negocios aplicada a la educación. Su propósito es descubrir patrones en los datos institucionales para analizar cuestiones útiles para la toma de decisiones estratégicas, orientando la gestión universitaria a partir de datos y modelos predictivos, siendo estos la prioridad que orienta el funcionamiento de la organización a nivel de gobernanza. Este enfoque busca predecir el desempeño, administrar el uso de recursos, detectar problemas potenciales y respaldar las operaciones administrativas. A diferencia de la analítica del aprendizaje, que se centra en los procesos formativos, la analítica académica amplía el análisis hacia indicadores claves de desempeño (KPI) que los sistemas educativos tradicionales no plantean de forma común (Lesjak et al., 2021; Mago & Khan, 2021; Nguyen, Gardner & Sheridan, 2020). En muchos casos, estas tecnologías se justifican como mecanismos para reducir costos, agilizar trámites y garantizar objetividad en la toma de decisiones, bajo la promesa de eliminar sesgos y mejorar tanto la eficiencia operativa como la reputación institucional (Marchena & de la Vega, 2021).

Desde el transhumanismo, la incorporación de la IA en la gestión universitaria sigue el ideal de perfeccionamiento tanto de la comunidad universitaria como de la organización mediante el control predictivo. Este arreglo se comparte con la analítica del aprendizaje, en el sentido de que todo proceso institucional puede mejorarse si es cuantificado. A partir de ello, la universidad se concibe como un sistema mediado tecnológicamente de manera que es permisible anticipar comportamientos y corregir desviaciones a partir del análisis de datos. Los algoritmos, al apoyar la toma de decisiones, sigue la promesa de la objetividad y la eliminación del error humano. Sin embargo, esta lógica tecnocientífica corre el riesgo de reducir la complejidad educativa a un conjunto de indicadores de desempeño y métricas.

Desde el poshumanismo, la IA en la gestión de las instituciones educativas no sustituye al ser humano, sino que se configura como un ensamblaje sociotécnico. Se puede pensar en la intervención de los algoritmos de forma material y simbólica en la conformación de formas híbridas de gobernanza entre sistemas automatizados y agentes humanos. Como explica Lamers et al. (2024), los algoritmos predictivos pronostican resultados a partir de datos pasados o en tiempo real, mientras que los predictivos prescriptivos sugieren cursos de acción mediante simulaciones y análisis de escenarios. Sin embargo, de manera conjunta, los algoritmos afectan las prácticas, orientan decisiones y generan significados. Quienes se ven implicados en estas dinámicas suelen incluso antropomorfizar a las algoritmos atribuyéndoles intenciones o juicios humanos para comprender su funcionamiento.

Implicaciones críticas de la IA en la universidad

Tras el abordaje conceptual trazando diferencias entre la mirada transhumanista y poshumanista en el campo de la educación superior, es relevante indagar en las implicaciones críticas y políticas que acompañan la expansión de la IA en la universidad. Estas transformaciones no solo inciden en el proceso de enseñanza aprendizaje, sino que se encuentran en todos los niveles analíticos desde lo micro hasta lo macro, dando lugar a un cambio en los fundamentos éticos, epistémicos y políticos de la educación como derecho universal y bien público. En este sentido se proponen tres ejes para orientar la discusión, ética, autonomía y mercantilización de la educación superior.

En primer lugar, la dimensión ética remite al desafío de mantener el sentido formativo y humano de la educación, aún en un entorno cada vez más automatizado y mediado por datos. Uno de los principales desafíos concierne a la protección de datos personales. El tratamiento masivo de información recopilada de la comunidad académica plantea riesgos asociados a la seguridad, la privacidad y uso posterior de los datos. Se reconoce que las instituciones de educación superior deben fortalecer los procedimientos de resguardo y gestión de datos, además de que es necesario un análisis continuo sobre las implicaciones presentes y futuras del uso de datos históricos en nuevos contextos. De manera general, se debe responder a través de políticas

universitarias que aseguren la transparencia, la equidad, el derecho de las partes interesadas de conocer y decidir sobre el uso de su información, así como la rendición de cuentas institucional (Alfy et al., 2019; Al-Zaharani & Alasmar, 2024; Cobo & Aguerrebere, 2018; Lesjak et al., 2021; Mago & Khan, 2021; Prinsloo & Slade, 2016; Webber & Zheng, 2024).

Junto con la protección de datos, la dimensión ética también debe atender las asimetrías de poder que se generan entre las IES y el alumnado. En muchos casos, las decisiones sobre la adopción de tecnologías se toman desde niveles administrativos o técnicos, sin una participación informada sobre cómo serán utilizadas. Esta relación desigual convierte al estudiantado en el grupo más vulnerable ante la implementación de sistemas de analítica y control predictivo. En este caso, la vulnerabilidad no solo refleja asimetrías de poder, sino que puede reproducir desigualdades preexistentes exponiendo a ciertos grupos al riesgo de exclusión o sesgo algorítmico (Prinsloo & Slade, 2016).

En este escenario, surgen preocupaciones relacionadas con la elaboración de perfiles y la segregación digital, que puede derivar en la exclusión de determinados grupos al acceso al conocimiento y la información. Esto se debe a que las arquitecturas tecnológicas que sustentan la IA no son neutras, sino que reproducen las condiciones históricas y sociales bajo las que fueron diseñadas, incorporando sesgos de máquina codificados en los conjuntos de datos utilizados para entrenar los algoritmos, bajo la posibilidad de reproducir patrones previos de exclusión y riesgo. De forma complementaria, la IA debe analizarse dentro del contexto del crecimiento de la tecnología educativa comercial y la influencia de diversos actores en el campo universitario, cuyas decisiones impactan en la distribución del conocimiento y median las oportunidades de participación. En este entendido, la importancia radica en desarrollar una lectura crítica de los datos, garantizando que las tecnologías no amplíen la brecha digital ni profundicen las inequidades existentes (Jarke & Breiter, 2019; George & Wooden, 2023; Webber & Zheng, 2024; Williamson et al., 2020; Williamson & Eynon, 2020).

En segundo lugar, términos de la autonomía universitaria, esta ha constituido de manera histórica un elemento fundamental en las universidades latinoamericanas, asociada a la autodeterminación, el autogobierno y la autogestión sin injerencias externas, siendo un proyecto emancipador en construcción, producto de las luchas sociales, la defensa del territorio y los bienes comunes frente al despojo capitalista (Oliveira & Modonesi, 2024). Sin embargo, la introducción de sistemas algorítmicos genera tensión con este principio, al trasladar parte del poder institucional de decisión hacia lógicas externas de tipo comercial por sobre lo académico. La adopción de modelos de gobernanza algorítmica presentes en plataformas o sistemas de analítica académica desarrollados por corporaciones tecnológicas implica que los procesos académicos y administrativos de las IES se regulen mediante infraestructuras con un diseño normativo e ideológico externo que ejerce control predictivo. Esto se vincula a la agencia performativa de las tecnologías digitales al replantear la experiencia educativa y reconfigurar ideas sobre el aprendizaje y la educación, incidiendo en los procesos de subjetivación (Amiel et al., 2022; Grimaldi & Ball, 2021; Decuyper, Grimaldi & Landri, 2021).

Por último, la IA se inserta en un contexto universitario marcado por la mercantilización. Diversos autores (Lesjak et al., 2021; Mago & Khan, 2021; Estrada-Real & Cantú-Ortiz, 2022) coinciden en que estas tecnologías participan en una dinámica de creciente mercantilización donde el valor de los datos adquiere un sentido económico y político, asociado con desigualdades en el poder y visibilidad entre naciones, comunidades científicas y grupos sociales. Cada interacción con tecnologías de aprendizaje deja rastros digitales que se convierten en la materia prima que da forma al modelo de negocio. Diversos participantes del sistema educativo son conscientes de esta maquinaria, emprendiendo la incorporación de tecnologías para mejorar su posicionamiento y reputación institucional. La educación superior adopta así la lógica de mercado en la que el alumnado se concibe como clientes, los programas académicos como productos y la universidad como una organización que busca maximizar su eficiencia. La gobernanza algorítmica refuerza esta dinámica, al trasladar decisiones

académicas y administrativas hacia infraestructuras de datos privadas diseñadas bajo la lógica de mercado. Williamson (2020) menciona respecto del proceso de mercantilización, que los proveedores privados de tecnologías digitales se han expandido en los sistemas de educación superior configurando una industria global que concentra poder económico para los corporativos.

En suma, estas tres aristas constituyen dimensiones interdependientes que revelan las tensiones del uso de la IA en la educación superior. Las cuestiones éticas se encuentran en intersección con el control algorítmico; la autonomía se redefine ante la intervención de la gobernanza algorítmica y la mercantilización plantea un horizonte económico que enmarca los procesos de transformación. En el contexto latinoamericano, donde la educación superior mantiene un sentido público producto de luchas históricas, estas implicaciones requieren de una mirada crítica y situada.

Pensar la inteligencia artificial desde América Latina

Pensar la IA desde América Latina implica reconocer que su despliegue ocurre en un contexto marcado por desigualdades históricas, dependencias tecnológicas y tensiones políticas. La región combina sistemas de educación superior heterogéneos, donde si bien, se comparten algunos aspectos en común de manera histórica y cultural, no se puede homologar su caracterización. La implementación de iniciativas de IA en la región se encuentra en una etapa temprana. Una explicación es que aspectos como la analítica de datos, requiere de factores técnicos, así como de condiciones organizativas, éticas, políticas y culturales, lo que representa un desafío estructural (Marchena y de la Vega, 2021). En este sentido, sólo las instituciones con acceso mayores recursos pueden implementar este tipo de iniciativas, mientras que otras permanecen al margen (Lesjak et al., 2021).

Por otra parte, la integración de tecnologías diseñadas por corporaciones externas a la universidad exige un enfoque participativo y crítico, en defensa de la autonomía universitaria que visibilice las desigualdades estructurales presentes en los sistemas educativos y permita deliberar sobre el papel de los actores involucrados en su diseño y uso

(William & Eynon, 2020). De igual manera, aunque la analítica de datos puede apoyar procesos académicos y administrativos, su aprovechamiento requiere de una humanización de los datos, no sólo en términos de optimizar el rendimiento, sino de identificar vulnerabilidades, reducir brechas y fortalecer la inclusión (Foster & Francis, 2020).

Esta idea es consistente con la importación de narrativas transhumanistas que asocian la IA con progreso, modernización y mejora institucional. No obstante, dichas narrativas no siempre dialogan con las realidades sociales, económicas, políticas y culturales de la región. El riesgo es que la IA se inserte como una tecnología descontextualizada, reforzando la lógica de mercado más allá de atender aspectos contextuales, políticos, culturales y políticos (Gašević, 2018). La propuesta es que en lugar de reproducir constructos extractivistas alineados a la datificación, las IES pueden reapropiar la IA como una herramienta para la colaboración, proponiendo alternativas que respondan a las necesidades y valores locales.

Este horizonte emancipador implica regular el papel de las tecnologías y promover marcos éticos propios, elaborados por las comunidades culturales de manera sensible a su contexto. Esto permite construir una visión de la IA de manera plural y situada donde se democratice el acceso al conocimiento, se garantice la equidad y se amplíen las formas de subjetivación en contextos atravesados por la desigualdad. Asimismo, es posible resistir a la lógica de mercado mediante el desarrollo de proyectos de IA de código abierto, redes interuniversitarias y plataformas públicas de gestión de datos que fortalezcan la soberanía tecnológica y reduzcan los costos. Esta propuesta es coherente con políticas de ciencia abierta, repositorios de datos colaborativos y recursos educativos compartidos (Amiel et al., 2022).

Conclusiones

La incorporación de la IA en la educación superior no constituye únicamente una innovación tecnológica, sino que implica una reconfiguración estructural de la universidad contemporánea. En este trabajo se ha analizado que la IA entendida desde las tensiones entre transhumanismo y poshumanismo, opera

simultáneamente como herramienta de optimización y como dispositivo transformación sociotécnica. Mientras el enfoque transhumanista privilegia la eficiencia, la mejora del rendimiento y la expansión de las capacidades humanas a partir de elementos artificiales, el poshumanismo reconoce la interdependencia entre humanos, máquinas y entornos.

El análisis de las dimensiones subjetivación, control predictivo y gobernanza algorítmica, permite comprender cómo la IA reconfigura las relaciones de poder, las formas de entretener el proceso de enseñanza aprendizaje y la toma de decisiones académica y administrativa. En este entramado, emergen tensiones éticas asociadas a la privacidad y al justicia de datos; tensiones políticas vinculadas a la autonomía universitaria y la soberanía tecnológica; y tensiones económicas derivadas de la mercantilización del conocimiento la datificación y el capitalismo de plataformas. Estas dinámicas no actúan de forma aislada, sino que conforman una racionalidad educativa orientada por la cuantificación, la predicción y la competencia, que redefine aspectos materiales y simbólicos de la educación superior.

Desde el contexto latinoamericano, estas tensiones adquieren un sentido particular. Las desigualdades estructurales y la presión por adoptar narrativas globales de innovación plantean desafíos específicos, pero también oportunidades para la reapropiación crítica de la tecnología. Pensar en la IA desde América Latina implica situarla desde las luchas históricas por la autonomía, la equidad y el derecho a la educación, reconociendo que las universidades pueden desempeñar un papel activo en la definición colectiva del rumbo tecnológico, de manera que la IA se conciba como una herramienta encaminada al servicio de la educación y la preservación de las funciones sustantivas.

La universidad se encuentra en una encrucijada, en este caso, ante reproducir el control algorítmico y la lógica de mercado que se refuerza con la IA, o redefinir su papel como espacio de emancipación y participación colectiva. En el segundo caso, es necesario una comprensión crítica de la tecnología, en la que se integre su potencial con los valores democráticos, en

defensa de la autonomía y del derecho universal a la educación, valores que han dado sentido a la educación superior en América Latina.

Referencias

- Alfy, S., Gómez, J. M., & Dani, A. (2019). Exploring the benefits and challenges of learning analytics in higher education institutions: A systematic literature review. *Information Discovery and Delivery*, 47(1), 25–34. <https://doi.org/10.1108/IDD-06-2018-0018>
- Al-Zaharani, A. M., & Alasmari, T. M. (2024). Exploring the impact of artificial intelligence on higher education: The dynamics of ethical, social and educational implications. *Humanities and Social Sciences Communications*, 11(912), 1–13. <https://doi.org/10.1057/s41599-024-03432-4>
- Amiel, T., Saraiva, F., Cruz, L. R. da, & Gonsales, P. (2022). Mapping surveillance capitalism in South American higher education. *Revista Latinoamericana de Tecnología Educativa*, 22(1), 221–239. <https://doi.org/10.17398/1695-288X.22.1.221>
- Chursinova, O., & Sinelnikova, M. (2024). Post-human and trans-human in the future perspective of the humanity. *Filosofija. Sociologija*, 35(1), 23–30. <https://doi.org/10.6001/fil-soc.2024.35.1.4>
- Cobo, C., & Aguerrebere, C. (2018). Building capacity for learning analytics in Latin America. En C. P. Lim & V. L. Tinio (Eds.), *Learning analytics for the global south* (pp. 58–67). Foundation for Information Technology Education and Development. <https://redi.anii.org.uy/jspui/handle/20.500.12381/307>
- Decuypere, M., Grimaldi, E., & Landri, P. (2021). Introduction: Critical studies of digital education platforms. *Critical Studies in Education*, 62(1), 1–16. <https://doi.org/10.1080/17508487.2020.1866050>
- Estrada-Real, A. C., & Cantú-Ortiz, F. J. (2022). A data analytics approach for university competitiveness: The QS world university rankings. *International Journal on Interactive Design and Manufacturing*, 16, 871–891. <https://doi.org/10.1007/s12008-022-00966-2>
- Foster, C., & Francis, P. (2020). A systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes. *Assessment & Evaluation in Higher Education*, 45(6), 822–841. <https://doi.org/10.1080/02602938.2019.1696945>
- Gašević, D. (2018). Include us all! Directions for adoption of learning analytics in the global south. En C. P. Lim & V. L. Tinio (Eds.), *Learning analytics for the global south* (pp. 3–21). Foundation for Information Technology Education and Development. <https://redi.anii.org.uy/jspui/handle/20.500.12381/307>

- Gherardi, S. (2021). A posthumanist epistemology of practice. En C. Neesham & S. Segal (Eds.), *Handbook of philosophy of management* (pp. 1–22). Springer. https://doi.org/10.1007/978-3-319-48352-8_53-1
- Gherardi, S., de Vaujany, F.-X., & Silva, P. (2024). Introduction: Too-human? Inquiring in-between different disciplinary areas in managing and organizing. En F.-X. de Vaujany, S. Gherardi, & P. Silva (Eds.), *Organization studies and posthumanism: Towards a more-than-human world* (pp. 15–34). Routledge. <https://doi.org/10.4324/9781032617169-1>
- Grimaldi, E., & Ball, S. J. (2021). Paradoxes of freedom: An archaeological analysis of educational online platform interfaces. *Critical Studies in Education*, 62(1), 114–129. <https://doi.org/10.1080/17508487.2020.1861043>
- Herbrechter, S. (2021a). Critical posthumanism, again. *New Horizons in the Humanities*, 1(1), 66–68. <https://doi.org/10.26522/POSTHUMANISMJOURNAL.V1I1.2585>
- Herbrechter, S. (2021b). *Critical posthumanism – An overview*. Heidelberg University. <https://stefanherbrechter.com/wp-content/uploads/2021/03/Critical-Posthumanism-An-Overview.pdf>
- Hong, S. (2024). Transcendence up for sale: Cracking the ontoexistential codes for Übermensch. *Consumption Markets & Culture*, 27(2), 152–177. <https://doi.org/10.1080/10253866.2023.2234293>
- Jarke, J., & Breiter, A. (2019). Editorial: The datafication of education. *Learning, Media and Technology*, 44(1), 1–6. <https://doi.org/10.1080/17439884.2019.1573833>
- Krüger, O. (2021). “The Singularity is near!” Visions of artificial intelligence in posthumanism and transhumanism. *International Journal of Interactive Multimedia and Artificial Intelligence*, 7(1), 16–23. <https://doi.org/10.9781/ijimai.2021.07.004>
- Lamers, L., Meijerink, J., & Rettagliata, G. (2024). Blinded by “algo economicus”: Reflecting on the assumptions of algorithmic management research to move forward. *Human Resource Management*, 63(2), 413–426. <https://doi.org/10.1002/hrm.22204>
- Lesjak, D., Natek, S., & Kohun, F. (2021). Big data analytics in higher education. *Issues in Information Systems*, 22(4), 320–333. https://doi.org/10.48009/4_iis_2021_346-359
- Mago, B., & Khan, N. (2021). A proposed framework for Big Data analytics in higher education. *International Journal of Advanced Computer Science and Applications*, 12(7), 685–691. <http://dx.doi.org/10.14569/IJACSA.2021.0120778>
- Marchena, S. G. F., & de la Vega, I. (2021). Adoption of big data analytics and its impact on organizational performance in higher education mediated by knowledge management. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(4), Article 221. <https://doi.org/10.3390/joitmc7040221>
- Merzlyakov, S. S. (2022). Posthumanism vs. transhumanism: From the “end of exceptionalism” to “technological humanism.” *Herald of the Russian Academy*

- of Sciences, 92(Suppl. 6), S475–S482. <https://doi.org/10.1134/S1019331622120073>
- Nassoura, A. B. (2022). Applied artificial intelligence applications in higher education institutions: A systematic review. *Webology*, 19(3), 1168–1183. <https://www.webology.org/abstract.php?id=2831>
- Negro, M. (2025). The surprising convergence of transhumanism and posthumanism. *Argumenta*, 2025, 1–17. <https://doi.org/10.14275/2465-2334/20250.neg>
- Nguyen, A., Gardner, L., & Sheridan, D. (2020). Data analytics in higher education: An integrated view. *Journal of Information Systems Education*, 31(1), 61–71. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3949946
- Nyathi, L. (2024). Unveiling the right to self-transformation and self-enhancement: Exploring transhumanism and its impact on human rights and the future of humanity. *Deusto Journal of Human Rights*, 13, 119–147. <https://doi.org/10.18543/djhr.3072>
- Oliveira, G., & Modonesi, M. (2024). Independence and emancipation: Latin American theorizations on the concept of autonomy. *Latin American Perspectives*, 51(4), 25–42. <https://doi.org/10.1177/0094582X241296816>
- Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182. <http://dx.doi.org/10.18608/jla.2016.31.10>
- Roden, D. (2014). Humanism, transhumanism and posthumanism. En D. Roden, *Posthuman life: Philosophy at the edge of the human*. Routledge. <https://doi.org/10.4324/9781315744506-2>
- Sapénko, R., & Trocha, B. (2020). Philosophical and cultural aspects of transhumanism and posthumanism. *Future Human Image*, 14, 64–72. <https://doi.org/10.29202/fhi/14/8>
- Selwyn, N., & Gašević, D. (2020). The datafication of higher education: Discussing the promises and problems. *Teaching in Higher Education*, 25(4), 527–540. <https://doi.org/10.1080/13562517.2019.1689388>
- Tecnológico de Monterrey. (2024). *Lineamientos para el uso ético de inteligencia artificial: Dirigido al profesorado del Tecnológico de Monterrey*. <https://tec.mx/sites/default/files/repositorio/integridad-academica/lineamientos-ia-profesores-tec-de-monterrey.pdf>
- UNESCO IESALC. (2023). *Oportunidades y desafíos de la era de la inteligencia artificial para la educación superior: Una introducción para los actores de la educación superior*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386670_spa
- University of Edinburgh. (2024). *Learning analytics: Principles and purposes, policy and procedures, and governance*. Registry Services – Academic Quality and Standards. <https://registryservices.ed.ac.uk/academic-services/policies-regulations/learning-and-assessment/learning-analytics>

- University of Michigan. (2025). *Learning analytics guiding principles*. Center for Academic Innovation. <https://ai.umich.edu/learning-analytics-guiding-principles/>
- Webber, K. L., & Zheng, H. Y. (2024). Artificial intelligence and advanced data analytics: Implications for higher education. *New Directions for Higher Education*, 204(207), 5–13. <https://doi.org/10.1002/he.20508>
- Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education*. Advance online publication. <https://doi.org/10.1080/17508487.2019.1691030>
- Williamson, B. (2020). Making markets through digital platforms: Pearson, edubusiness, and the (e)valuation of higher education. *Critical Studies in Education*, 63(4), 444–460. <https://doi.org/10.1080/17508487.2020.1737556>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>
- Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives. *Teaching in Higher Education*, 25(4), 351–365. <https://doi.org/10.1080/13562517.2020.1748811>
- Williamson, B., Macgilchrist, F., & Potter, J. (2023). Re-examining AI, automation and datafication in education. *Learning, Media and Technology*, 48(1), 1–5. <https://doi.org/10.1080/17439884.2023.2167830>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>

Espesor perceptivo: Una aproximación al paisaje algorítmico desde la teoría del arte

Perceptual thickness: An approach to algorithmic landscape from art theory

Constanza Rivano Delzo

constanza.rivano@uchile.cl

*Centro Interdisciplinario de Estudios en Filosofía, Artes y Humanidades de la
Universidad de Chile*

ORCID ID 0009-0009-6380-6152

Resumen: Este artículo propone una articulación filosófica entre la ontología del arte de Martin Heidegger y la filosofía del paisaje de Georg Simmel para pensar el estatuto estético del paisaje generado por inteligencia artificial. En Heidegger, la obra de arte es el lugar donde acontece la verdad como desocultamiento. Tal acontecer se sostiene en una pugna constitutiva entre mundo, horizonte histórico de sentido que orienta lo significativo y, tierra, dimensión de reserva que confiere espesor al aparecer. En Simmel, el paisaje surge cuando la mirada cultural enmarca la infinitud de la naturaleza y la configura en una totalidad perceptiva; el paisaje es una forma y, a la vez, una operación cultural de unificación que preserva un resto de multiplicidad. La hipótesis central del artículo sugiere que, en el arte de IA, esta doble estructura de apertura y reserva se mantiene, aunque desplazada hacia una fenomenología desmaterializada.

Para fundamentar esta tesis se introduce la categoría de “espesor perceptivo”: una experiencia de densidad sensible que emerge cuando el mundo se abre como reconocimiento cultural y, al mismo tiempo, algo resiste en la percepción. En la pintura tradicional, esa resistencia se anuda a la materialidad (pigmento, gesto, soporte); en las imágenes algorítmicas, la resistencia adopta la forma de una ilusión de naturalidad que se ofrece y se retira al mismo tiempo. El espectador reconoce motivos y gramáticas del paisaje heredadas por la historia del arte, pero percibe la

pérdida de suelo, el desanclaje de una naturaleza sin pertenencia. El espesor, así entendido, deja de depender de la materia y pasa a concentrarse en la fricción entre memoria cultural y cálculo técnico.

El estudio de caso de *Landscape Paintings* (2014–2017) de Quayola permite verificar este desplazamiento. La serie se construye a partir de registros audiovisuales y procedimientos algorítmicos que sintetizan color, luz y estructura en imágenes que oscilan entre lo reconocible y lo abstracto. Se produce una apertura de mundo por evocación de la tradición paisajística (composición atmosférica, horizontes, ritmos cromáticos) y, simultáneamente, un espesor que aparece como resto fenomenológico: transiciones demasiado perfectas, vibraciones lumínicas, regularidades que generan extrañamiento. Ese resto opera como tierra en sentido heideggeriano, aunque ya no se apoye en materialidad visible. Lo que salta al frente es la verdad de una mediación técnica que funda experiencia: el paisaje persiste como categoría de verdad, ahora traducida al régimen algorítmico.

De este modo, el artículo sostiene que el paisaje de IA no clausura el marco heideggeriano ni el simmeliano; más bien los reubica. Mundo y tierra continúan en pugna, pero el espesor que los vincula se vuelve perceptivo antes que material. El resultado es una poética del reconocimiento y la pérdida: la belleza aparece cuando la memoria del paisaje se enfrenta a su propia desmaterialización y el espectador se afirma en la densidad de ese intervalo.

Palabras clave: Heidegger; Simmel; paisaje algorítmico; Espesor Perceptivo; Quayola.

Abstract: This article proposes a philosophical articulation between Martin Heidegger's ontology of art and Georg Simmel's philosophy of landscape to reassess the aesthetic status of AI-generated landscape. In Heidegger, the artwork is the site where truth occurs as unconcealment. Such occurrence rests on a constitutive strife between world, the historical horizon of meaning that orients significance and, earth, the dimension of reserve that lends depthness to appearance. In Simmel, landscape emerges when cultural perception frames the infinitude of nature and configures it as a perceptual whole; landscape is both a form and a cultural operation of unification that preserves a remainder of multiplicity. The article's central claim is that, in AI art, this dual structure of opening and reserve persists, yet it is displaced toward a dematerialized phenomenology.

To conclude this claim, the article introduces the category of perceptual depthness: a felt density that arises when world opens as cultural recognition and something simultaneously resists within perception. In traditional painting, resistance adheres to materiality (pigment, gesture, support). In algorithmic images, resistance takes the form of an illusion of naturalness that both offers itself and withdraws. The viewer recognizes inherited motives and grammars of landscape from art history, while sensing the loss of ground, the unmooring of a nature without belonging. Perceptual depthness thus ceases to rely on matter and concentrates in the friction between cultural memory and technical computation.

The case study of Quayola's *Landscape Paintings* (2014–2017) substantiates this shift. The series builds on audiovisual captures and algorithmic procedures that synthesize color, light, and structure into images hovering between recognizability and abstraction. A world is opened through the evocation of the landscape tradition (atmospheric composition, horizons, chromatic rhythms) while depthness appears as a phenomenological remainder: overly perfect transitions, luminous tremors, regularities that induce estrangement. This remainder functions as earth in the Heideggerian sense, although no longer grounded in visible materiality. What comes to the fore is the truth of a technical mediation that founds experience: landscape persists as a category of truth, now translated into the algorithmic regime.

Accordingly, the article argues that AI landscape neither cancels nor repeats Heidegger's and Simmel's frameworks; rather, it relocates them. World and earth remain in strife, yet the depthness binding them becomes perceptual rather than material. The outcome is a poetics of recognition and loss: beauty arises when the memory of landscape confronts its own dematerialization and the viewer steadies themselves in the density of that interval.

Keywords: Heidegger; Simmel; algorithmic landscape; Perceptual Depth; Quayola.

Introducción

Este estudio propone una lectura ontológica del arte digital contemporáneo a partir del cruce entre *El origen de la obra de arte* de Martin Heidegger y la *filosofía del paisaje* de Georg Simmel. En Heidegger, la obra acontece como verdad en acto, un desocultamiento que se sostiene en la tensión entre *mundo*;

entendido como horizonte de sentido compartido y, *tierra*, concebida como una reserva que da espesor a lo visible (Heidegger, 2002). Si bien esta articulación enfatiza el carácter ontológico del aparecer, Simmel ofrece un complemento decisivo para comprender su dimensión perceptiva. En su análisis, el paisaje emerge cuando la mirada cultural organiza la multiplicidad en una totalidad perceptiva, donde forma y fondo coexisten en equilibrio dinámico (Simmel, 2013). De este modo, ambas perspectivas convergen en una tesis común: el arte instituye modos de aparecer que orientan una comunidad histórica.

La irrupción del arte generado por inteligencia artificial reubica esa tensión. Sistemas algorítmicos producen imágenes que reactivan repertorios culturales mientras operan mediante cálculos opacos a la percepción (Manovich, 2018). La experiencia ya no se apoya en pigmento o trazo, pero conserva una fricción sensible. En el caso del paisaje como género pictórico en su forma digital, lo familiar convive con la ajenidad técnica de su génesis. Para nombrar esta condición propongo la noción de espesor perceptivo: una densidad fenomenológica que se experimenta cuando la apertura de mundo queda trenzada con una reserva que no es material, sino perceptible como límite, desfase o exceso dentro de la imagen (Donnarumma, 2023).

La serie *Landscape Paintings* de Jacopo Quayola ofrece un caso ejemplar. Sus paisajes algorítmicos convocan la memoria pictórica del género y, a la vez, hacen sentir la fricción entre cálculo y percepción. En ese rozamiento, la pugna heideggeriana se actualiza: el mundo se abre en la evocación cultural del paisaje; la tierra se insinúa como resto perceptivo que impide la clausura de la transparencia. El objetivo del artículo es, primero, precisar la relación mundo–tierra y la configuración forma–fondo en sus fuentes filosóficas; segundo, mostrar cómo *Landscape Paintings* permite formalizar el espesor perceptivo como categoría estética para pensar el paisaje en la era algorítmica; y, tercero, argumentar que esta categoría describe la continuidad del arte como ámbito de verdad en condiciones de desmaterialización técnica.

Heidegger: verdad, mundo y tierra en la obra

La intervención de Heidegger en estética parte de una sospecha; hablar de arte como asunto de gusto o contemplación empobrece aquello decisivo que acontece en la obra. En *El origen de la obra de arte*, desplaza el eje desde la subjetividad estética hacia la ontología del aparecer. En la obra, irrumpe la verdad como desocultamiento. Este desplazamiento no elimina la experiencia sensible; le otorga fundamento al situarla en el terreno donde algo llega a mostrarse como lo que es. De ahí que la pregunta por el origen de la obra no indague genealogías biográficas ni contextos sociales, sino el punto desde el cual una obra es obra: el *poner-en-obra* en función de su acontecer (Heidegger, 2002).

Una parte sustancial de la discusión sobre estética y filosofía del arte en Heidegger, ha mostrado que esta reorientación no constituye una estética negativa, sino una crítica de la estética moderna en favor de la fecundidad ontológica del arte (Thomson, 2012). En esa clave, la obra no adorna un mundo previo, sino que abre mundo, instituye modos de relevancia y otorga contornos históricos a lo significativo (Young, 2001). El gesto heideggeriano exige, por tanto, una lectura de la obra como acontecimiento de verdad, y no como simple representación. Esta tesis inaugura el campo en el que sus nociones de mundo y tierra cobran precisión, pues ambas nombran dimensiones del aparecer que se entrelazan en la obra y generan la tensión del aparecer de verdad sin reducirse a materias o ideas separadas (Thomson, 2012; Young, 2001).

El primer paso del ensayo explora el carácter de cosa (*Dinghaftigkeit*). Heidegger no se conforma con declarar que la obra tiene materialidad; investiga cómo ese *elemento de cosa* sostiene el aparecer. Frente a dos atajos; entendiendo cosa como suma de sensaciones o como portadora de propiedades útiles, propone pensar la cosa como aquello que resiste y, al resistir, permite que algo se muestre con estabilidad (Heidegger, 2002). Este análisis abre un acceso a la distinción entre útiles (equipamiento que se consume en el uso) y obra (aquello en lo que el aparecer se instala y se hace perdurable).

Comentadores han subrayado que aquí asoma el tránsito desde *ser-a-la-mano* y *estar-a-la-mano* de *Ser y tiempo* hacia una fenomenología del soportar que hace posible la escena del desocultamiento. En términos sistemáticos, Heidegger prepara el terreno para mostrar que la verdad no se agota en correspondencias proposicionales: exige un ámbito donde lo significativo obtenga figura y persista. Así, se comprende que el origen del arte sea entendido como posibilidad de instauración, es decir, aquello desde lo cual emergen artista y obra en copertenencia, antes que entidades autosuficientes (Heidegger, 2002).

Sobre ese trasfondo, mundo (*Welt*) nombra el conjunto articulado de referencias, fines y valores donde algo cuenta como valor. El mundo no suma objetos; delimita un horizonte histórico de inteligibilidad que confiere sentido a las cosas y orienta la praxis de una comunidad. Entonces, una obra abre mundo en la medida en que concentra y proyecta una configuración de significación que, al volverse sensible, ilumina a un nosotros histórico (Heidegger, 2002).

La literatura especializada ha insistido en esta función instituyente: la obra condensa un saber tácito compartido y lo retorna a la comunidad para su autocomprensión a través de la obra (Thomson, 2012). En términos hermenéuticos, la obra no ilustra un trasfondo previo; le da fisonomía, intensifica su perfil y lo lanza a futuro como posibilidad de comprensión. De ahí que ejemplos como el templo griego no interesen por su iconografía, sino por el modo en que configuran un campo de sentido donde dioses, mortales, festividades y orden del lugar se determinan mutuamente (Heidegger, 2002). Esta lectura, clásica en los estudios heideggerianos, ha sido articulada con detalle en debates de estética analítica y fenomenológica, que subrayan la función de la obra en la explicitación pública de un mundo.

La tierra (*Erde*), en cambio, no equivale a materia bruta ni a un depósito de recursos. Por su parte, la tierra designa la dimensión del aparecer que se reserva, se guarda y, al hacerlo, sostiene el mostrarse. Esta tierra nombra lo que, al cerrarse, preserva su esencial retraimiento y confiere a lo patente un espesor que lo sustrae a la transparencia total (Heidegger, 2002).

La recepción académica ha mostrado que esta categoría introduce una semántica de la opacidad constitutiva: la experiencia de verdad incluye siempre sombra, resto, exceso que no se deja domesticar sin pérdida (Colapinto, 2005; Thomson, 2012). Pensada desde la obra de arte, tierra es el modo en que el material (piedra, color, palabra, entre otros), se mantiene activo como resistencia y, por eso, nunca se agota en la función representativa. Dejar ser la tierra significa permitir que ese retraimiento acontezca en la obra como densidad sensible, como gravedad que guarda lo no dicho. Con ello, la verdad acontece como desocultamiento que exige lo velado como condición de posibilidad, y la obra es el lugar privilegiado para experimentar esa co-implicación de luz y resguardo (Colapinto, 2005; Heidegger, 2002).

Mundo y tierra no forman dos capas yuxtapuestas. Su relación es polémica y cooriginaria. Heidegger sostiene que es una relación de *pugna* (*Streit*) lo que define la esencia de la obra. Mundo busca abrir caminos y perfilar figuras de sentido; mientras tanto, la tierra retrae y preserva lo que le pertenece a su tiempo, lo que desborda toda determinación fuera de su propio contexto. Así, la obra sostiene la fricción, la hace visible y fecunda. Por esto, la experiencia ontológica en la obra de arte, se da como equilibrio inestable entre claridad orientadora y resistencia que afirma límites (Heidegger, 2002). Esta lectura permite comprender por qué el arte no se agota en mensajes descifrables, aquello que la obra pone al frente siempre vuelve a remitirse a una reserva que lo sostiene. La pugna es, entonces, la forma fenomenológica del desocultar que, al acontecer, vuelve patente su propia dependencia de lo que permanece en sombra. De ahí su valor crítico frente a estéticas de la transparencia y frente a instrumentaciones tecnológicas que persiguen visibilidad sin resto (Thomson, 2012).

Desde aquí se esclarece el sentido de verdad implicado. Verdad ya no es adecuación entre juicio y cosa; es el acontecer del desocultamiento que instituye un ámbito donde algo puede comparecer como lo que es. Este acontecer tiene estructura cuádruple: requiere un poner proyectivo (el obrar del artista), un instalar material (el trabajo de la cosa), un preservar por parte

de quienes acogen la obra, y una historia que se reconfigura al irrumpir esa figura en el tiempo (Heidegger, 2002).

En esta clave, la obra no dice una verdad externa; deja que verdad acontezca como escena compartida. Buena parte de la discusión en la filosofía del arte ha mostrado que este concepto de verdad evita psicologismos del gusto y semánticas puramente proposicionales, al tiempo que explica la fuerza exhibicional del arte para configurar horizontes públicos. El resultado es un concepto robusto de experiencia ontológica y como consecuencia, experimentar una obra de arte, es participar de un abrirse histórico de significados que reubican prácticas, creencias y cuerpos. La obra deviene, así, un foco de redistribución de relevancias, y su verdad consiste en ese redistribuir, no en decir algo sobre un mundo ya fijado.

Este marco se consolida cuando Heidegger analiza la función ejemplar de obras como el templo griego o los zapatos de Van Gogh. El interés de estos ejemplos no reside en adjudicarles contenidos simbólicos, sino en exhibir cómo la obra hace lugar. En el templo, el emplazamiento determina un ámbito de prácticas y creencias; en la pintura, la trama cromática y el gesto pictórico intensifican un mundo campesino que se deja entrever a partir de rastros (Heidegger, 2002). Incluso las controversias historiográficas han servido para precisar la tesis: lo discutible de una atribución concreta no invalida la idea de que la obra abre un campo de referencia en el que usos, fines y valoraciones se reordenan (Harries, 2009). En este punto, la crítica contemporánea ha subrayado la relevancia de la *fiabilidad* (*Zuverlässigkeit*) de los materiales: la obra confía en la resistencia de sus elementos para sostener el aparecer, y esa confianza delimita el régimen de experiencia que habilita (Colapinto, 2005). Así se entiende que la materialidad no se consuma ni desaparezca: deviene suelo que hace posible el despliegue del mundo. En esta codeterminación, el sentido de verdad recupera su espesor y el arte su potencia instituyente (Colapinto, 2005; Harries, 2009).

Esta tensión, Heidegger, la describe como una proyección poética. La obra abre posibilidades de comprensión que orientan porvenir sobre un suelo ya recibido (Heidegger, 2002). La investigación especializada lo ha enfatizado:

las grandes obras recapitulan un saber tácito de época y, al exhibirlas, son repensadas (Thomson, 2012; Young, 2001). De ahí que incluso cuando una obra pierde vigencia cultural, conserva su capacidad de reabrir lo significativo en nuevas constelaciones no desaparece, sino que cambia de régimen, como han defendido lecturas atentas a la continuidad entre ontología del arte y temporalidad histórica (Singh, 1990).

En esta perspectiva, el desocultamiento de la obra de arte puede entenderse como el instalarse de una posibilidad de mundo cuyo aparecer exige el resguardo de una tierra que ningún concepto agota. La verdad a la que alude Heidegger es, entonces, una praxis de apertura que instituye figuras históricas de sentido sosteniéndose en una reserva irreductible. Esa es la singular contribución del arte: fijar, en lo sensible y ontológico, la tensión viva entre abrir y guardar que mantiene habitable la experiencia común (Heidegger, 2002; Thomson, 2012).

Heidegger y la pregunta por el arte en la era algorítmica

El desarrollo del arte generado por inteligencia artificial desafía los marcos ontológicos que sustentaron la reflexión estética durante siglos. Este tipo de arte, conocido como *AI-generated art*, designa producciones visuales, sonoras o textuales en las que algoritmos de aprendizaje automático crean o co-crean resultados a partir de grandes conjuntos de datos. Lejos de constituir un mero instrumento, la IA interviene en la génesis misma de la obra, modificando la relación entre creación, materialidad y recepción. Manovich (2018) señala que la estética algorítmica ya no se define por el gesto del artista, sino por una lógica estadística que reconfigura patrones culturales previos. Donnarumma (2023) añade que la agencia tecnológica se vuelve constitutiva de la experiencia artística contemporánea. Desde este punto de vista el arte algorítmico no imita la creatividad humana, sino que produce zonas de indeterminación entre lo humano y lo maquínico.

Ante este escenario, la pregunta heideggeriana sobre la obra de arte, adquiere nueva urgencia. Si la obra es el lugar donde la verdad de lo humano y su relación con el tiempo y la historia se desocultan y acontecen, ¿qué tipo de verdad comparece cuando ese acontecer se realiza a través de un proceso

tecnológico que no posee conciencia, intención ni experiencia del ser? Para entender esto, el problema no está centrado en la capacidad creadora de esta nueva tecnología, es decir, pensar en concederle una agencia artística. Incluso, si así lo fuera ¿Cómo su modo de operar transforma la estructura del desocultamiento?, dicho de otro modo, ¿Cómo se ve adaptada la pugna entre mundo y tierra, que hace de la obra un acontecimiento de verdad?

La dificultad para pensar el arte generado por IA desde Heidegger radica en que el *Streit* (la pugna entre mundo y tierra) parece alterado por la sustitución del trabajo manual y material por la mediación algorítmica. El mundo que se abre en estas obras pertenece al repertorio cultural inscrito en los datos de entrenamiento, mientras que la tierra ya no remite a un soporte tangible, sino a un desconocimiento perceptivo propio del instrumento. Según Hansen (2020), el arte computacional produce un nuevo régimen fenomenológico donde la experiencia artística depende de la interacción entre percepción humana y procesos invisibles de cálculo. La obra generada por IA se sitúa, por tanto, en un umbral en donde el mundo se construye con imágenes heredadas, pero su tierra, aquello que, mencionamos, resiste y guarda, se constituye como límite inaccesible, es decir, como fricción entre la transparencia digital y la densidad sensible. Desde una lectura heideggeriana, podríamos decir que la pugna persiste, pero se ha desplazado, ¿dónde está? El arte algorítmico instala un nuevo tipo de tensión entre desocultamiento y reserva, donde la tierra ya no se experimenta en la materia física, sino en la aparición de un resto perceptivo que impide la clausura del sentido.

Dentro de los géneros pictóricos de la tradición artística, este artículo propone al paisaje como género pictórico por excelencia para analizar el desplazamiento de la pugna en sentido heideggeriano. El desafío ontológico obra de arte generadas con IA consiste en que el artista ya no se sitúa en relación inmediata con la tierra, sino mediado por una red de procesos tecnológicos que operan fuera de su control. En este sentido, la IA radicaliza la tendencia moderna a convertir el arte en producto de la técnica, pero, también abre un campo para repensar el carácter originario del arte como creación, como ese traer a la presencia lo que aún solo existía como posibilidad. Floridi (2019) sugiere que la IA debe comprenderse como una

agencia no intencional capaz de producir efectos informacionales en el mundo. Trasladado al ámbito de lo artístico, esto implica que la obra algorítmica no representa, sino que reconfigura las condiciones mismas de aparición de lo sensible.

Heidegger podría reconocer en este gesto una forma inédita del desocultamiento del mundo, ya que, el horizonte de sentido acontece a través de mediaciones tecnológicas que, al operar sin conciencia, evidencian el poder instaurador de la tecnología. En este sentido podríamos establecer que la categoría de mundo se mantiene a pesar de que ahora la IA forma parte del dinamismo de la obra. El peligro está en que, dada las condiciones de funcionamiento de la herramienta, la producción artística podría ser una obra sin resistencia. Es decir, sin el resguardo de la tierra en sentido heideggeriano. De ahí la importancia de buscar, dentro del arte de IA, los lugares donde la obra aún conserva la pugna. Encontrar ese lugar donde algo se guarda, se resiste y, en ese resguardarse, otorga profundidad ontológica a la imagen.

Georg Simmel y la aproximación al paisaje algorítmico

El pensamiento de Georg Simmel sobre el paisaje ofrece una vía complementaria para comprender cómo la experiencia ontológica de la obra de arte conserva su espesor incluso cuando la materialidad tradicional de la obra se desvanece. En *Filosofía del Paisaje* (2013), Simmel define el paisaje como una configuración perceptiva en la que la naturaleza deja de ser una continuidad inabarcable para volverse totalidad significativa. Este acto de configuración es una operación cultural, un proceso mediante el cual la mirada organiza lo disperso en forma coherente. Lo decisivo no es lo natural en sí, sino el modo en que lo natural sensible se vuelve visible como mundo. Este ejercicio de mediación expresado por Simmel es la que permite trasladar la filosofía simmeliana hacia una lectura contemporánea, en la que el paisaje generado por inteligencia artificial ya no está sostenido por pigmentos ni por el gesto del artista, pero sigue siendo una forma de tensión entre apertura y pérdida.

En la filosofía simmeliana, el paisaje es una forma de unidad que nunca anula la multiplicidad que la compone. Cada paisaje encierra la

presencia simultánea de orden y exceso. La mirada, como instrumento, logra aprehender un fragmento del mundo como totalidad, pero esa totalidad mantiene en reserva lo que excede su forma. Frisby (1992) señala que el pensamiento de Simmel no aspira a la clausura sino a la coexistencia dinámica entre forma y vida. Esta coexistencia anticipa lo que podemos entender como el espesor perceptivo.

El concepto de *Rahmung* (enmarcamiento) de Simmel ayuda a comprender esta nueva condición. El paisaje es un recorte cultural que delimita un fragmento del mundo para hacerlo visible. Esta operación de enmarcar (material o perceptiva) no sólo organiza la visión, sino que crea la ilusión de una totalidad armónica. Harrington (2004) destaca que este enmarque es siempre una forma de mediación: el paisaje es cultural incluso cuando parece natural. En el arte generado por inteligencia artificial, esa mediación adopta un nuevo rostro: el marco ya no es el lienzo o la ventana, sino la interfaz digital, la visualización estadística, el sistema que traduce datos en imagen. La operación cultural que antes realizaba el ojo del artista ahora la ejecuta el algoritmo, pero el efecto sigue siendo el mismo: la constitución de una totalidad perceptiva que nos sitúa frente a lo familiar. Este reconocimiento cultural (ese mundo que se abre ante nosotros) conserva la potencia de la verdad en la obra de arte, aunque su soporte ya no sea la materia sino la memoria.

Como fue mencionado anteriormente, en Heidegger, la obra acontece como pugna entre mundo y tierra: el mundo se abre como horizonte histórico, mientras que la tierra se resguarda como fondo que se retira. Esta tensión es el corazón del arte como desocultamiento. Pero cuando nos enfrentamos a un paisaje generado por IA, esta estructura se desestabiliza. El mundo aún comparece: el espectador reconoce la memoria cultural del paisaje, las convenciones pictóricas que evocan una historia visual compartida. Lo que falta, o se transforma radicalmente, es la tierra. Ya no hay pigmento, trazo ni materia visible. Tampoco hay rastro del artista, pues la IA, como señala Donnarumma, es una agencia que no deja huella, es una producción sin cuerpo (2023). La pregunta se impone: ¿qué es lo que resiste

cuando no hay materia? ¿Qué constituye ahora el espesor de la experiencia estética?

Lo que resiste en estas obras es la ilusión de lo natural. Esa ilusión opera como una nueva tierra, no en sentido material, sino fenomenológico. La imagen de paisaje generada por IA nos atrae por su familiaridad, por la evocación de un mundo que parece accesible y habitable. Pero esa familiaridad es ilusoria: el paisaje no pertenece a ningún lugar, no tiene suelo, ni atmósfera, ni tiempo. Se trata de una naturaleza descontextualizada, reconstruida por procedimientos estadísticos. Según plantea Hansen (2020), la estética digital suspende la materialidad del mundo como referencia y la sustituye por la presencia de lo calculado. El espectador, al contemplar este paisaje, experimenta simultáneamente cercanía y extrañamiento: es convocado por un mundo reconocible y, a la vez, confrontado con su pérdida. Ese desgarró entre la promesa de lo natural y la conciencia de su desaparición es lo que da forma al espesor perceptivo contemporáneo.

El espesor perceptivo, en este contexto, no depende de la resistencia física de la materia, sino del modo en que la percepción reconoce la fractura entre apariencia y acontecimiento. Allí donde el trazo pictórico revelaba la fricción perceptiva entre gesto y soporte, la imagen algorítmica revela la tensión entre memoria y cálculo. Lo que se retrae, es decir, aquello que sería la tierra en sentido heideggeriano, no es ya lo material, sino la condición de lo real su pertenencia a un mundo compartido. El arte algorítmico mantiene viva la pugna entre mundo y tierra en la forma de una nostalgia perceptiva. El mundo se abre en la evocación cultural del paisaje, mientras que la tierra se retira en el simulacro de su pérdida. El espectador experimenta el desocultamiento como melancolía: la verdad que acontece es la conciencia de un mundo que persiste solo como imagen.

Esta interpretación permite reconsiderar la vigencia del pensamiento de Simmel y Heidegger en la filosofía del arte contemporánea. En ambos casos, la experiencia estética se funda en una relación de tensión entre forma y fondo, entre mundo y tierra, pero hoy esa tensión se juega en un terreno desmaterializado. El paisaje algorítmico, como sugiere Manovich (2018), trae

a la presencia el imaginario del paisaje romántico y lo traduce en lenguaje de datos. Esta traducción no destruye la experiencia frente a la obra, sino que la transforma: el espesor perceptivo ya no proviene de la materia, sino de la distancia entre lo calculado y lo vivido, entre la ilusión de lo natural y su imposibilidad. En este desplazamiento se encuentra la nueva tierra del arte digital: una resistencia que no es física, sino perceptiva, una reserva de significado que se oculta en la familiaridad misma de la imagen.

La relación entre Simmel y Heidegger permite comprender este fenómeno con mayor profundidad. En Simmel, el paisaje mantiene su vitalidad gracias a la tensión entre unidad y multiplicidad. En Heidegger, la obra sostiene su verdad en la pugna entre apertura y reserva. En el arte de IA, ambas tensiones convergen: la unidad perceptiva del paisaje se abre como mundo, mientras que su pérdida de contexto, su naturaleza desanclada, artificial, actúa como tierra que resiste. El espesor perceptivo, concepto inspirado en el peso de la memoria, es, por tanto, la forma contemporánea de la pugna heideggeriana. Es la experiencia sensible de la verdad en una época en que la materia se ha vuelto invisible y la huella del artista se ha disuelto. Lo que salta al frente no es ya el trazo o el pigmento, sino la memoria de lo natural convertida en simulacro.

Este desplazamiento tiene implicaciones hermenéuticas decisivas. Si en Heidegger la obra de arte abre un mundo en el que una comunidad se reconoce históricamente, el arte algorítmico abre un mundo de reconocimiento sin pertenencia. La imagen generada por IA conserva la estructura de sentido (el horizonte cultural del paisaje), pero carece del suelo histórico que la sustentaba. La verdad que se pone en obra es la del desarraigo: un mundo sin tierra. En esa pérdida, se revela algo esencial, la persistencia de la experiencia estética como anhelo de espesor. El espectador busca en la imagen una densidad que ya no está en la materia, sino en el gesto de reconocimiento. La tierra se vuelve recuerdo; su resistencia consiste en recordarnos que lo natural, tal como lo concebíamos, ha sido desplazado por la tecnología.

El espesor perceptivo, entonces, es la categoría que permite pensar la continuidad entre la estética del paisaje y la ontología del arte en la era digital. Designa la experiencia en la que la verdad acontece no por la presencia física de la obra, sino por la conciencia de su pérdida. Es el residuo perceptivo de la tierra en la imagen desmaterializada, el resto sensible que mantiene viva la pugna entre mundo y ocultamiento. Desde esta perspectiva, el paisaje generado por IA no representa una ruptura con la tradición, sino su metamorfosis: el arte sigue siendo el ámbito donde lo visible se enfrenta a su fondo invisible, donde la memoria de lo natural se resiste a desaparecer completamente en la transparencia de la técnica.

Landscape Paintings – Quayola

Jacopo Quayola (Roma, 1982) se ha consolidado como una de las figuras más relevantes del arte contemporáneo que explora las intersecciones entre tecnología, memoria visual y estética clásica. Su obra se inscribe en un linaje de artistas que, lejos de concebir la inteligencia artificial como herramienta instrumental, la entienden como medio ontológico de exploración de la imagen. En este sentido, Quayola no produce imágenes digitales que imiten el arte tradicional, sino que se sitúa dentro de una genealogía de pensamiento visual que interroga las condiciones mismas de lo pictórico. Este enfoque sitúa su práctica en el corazón del debate filosófico contemporáneo sobre la agencia estética de la inteligencia artificial.

Quayola se formó en Londres, donde desarrolló un interés por las tensiones entre arte e ingeniería, y comenzó a trabajar con algoritmos generativos, proyecciones inmersivas y sistemas de datos. Su obra, que abarca video, instalación, escultura y software, se caracteriza por explorar los límites entre representación y abstracción, entre el orden geométrico y el caos perceptual. En este sentido, Quayola reinterpreta los géneros históricos de la pintura (el retrato, el paisaje, la naturaleza muerta) como sistemas de observación y cálculo. Su serie *Landscape Paintings* (2014–2017) representa el punto culminante de esta búsqueda: un proyecto que, bajo la apariencia de continuidad con la tradición pictórica, reformula la experiencia del paisaje en clave algorítmica.

Landscape Paintings consiste en una serie de imágenes generadas a partir de grabaciones de video en alta definición de paisajes reales (bosques, montañas, valles) que luego son procesadas mediante programas de análisis y síntesis visual. El algoritmo traduce los datos cromáticos, lumínicos y espaciales en formas geométricas que, al acumularse, generan una imagen fluctuante, entre lo reconocible y lo abstracto. El resultado son composiciones que evocan a Turner o Cézanne, pero en las que la materia pictórica ha sido reemplazada por campos de color pixelado y planos fractales. Su interés radica en revisitar la idea tradicional de paisaje a través del lenguaje de la computación en donde el proceso de ver es repensado debido a los algoritmos. La pintura se convierte así en evento de percepción y cálculo, donde la máquina asume el papel de mediadora entre percepción y el mundo.

Desde la perspectiva de la teoría del arte, *Landscape Paintings* es una reflexión sobre la posibilidad misma del paisaje como categoría estética en la era tecnológica. Si, como planteaba Simmel (2013), el paisaje surge cuando la naturaleza es aprehendida como totalidad unificada por la mirada, entonces el gesto algorítmico de Quayola repite esa operación en otro plano: el algoritmo observa, segmenta y recompone la naturaleza en una forma total. El paisaje es nuevamente resultado de una mediación cultural, pero ahora la mediación es pensada artificialmente.

En palabras de Manovich (2018), este tipo de obras marcan la transición hacia un estadio *postfotográfico* donde las imágenes no documentan el mundo, sino que lo simulan a partir de datos, incluso acercándose más al pensamiento pictórico que al fotográfico. Quayola no pinta un bosque: programa un sistema para que la máquina imagine uno. Sin embargo, la experiencia estética sigue presente. El espectador percibe una unidad que se abre como mundo, es decir, un horizonte de sentido que remite a la memoria cultural del paisaje y simultáneamente experimenta el extrañamiento de enfrentarse a una naturaleza sin suelo, generada por cálculo.

Este desplazamiento resuena con la lectura heideggeriana del arte como acontecimiento de verdad. En *El origen de la obra de arte*, Heidegger

(2002) define la obra como el lugar donde la verdad del ser acontece en la pugna entre mundo y tierra. En los paisajes de Quayola, el mundo comparece en la evocación cultural de la pintura paisajística: los tonos dorados de un atardecer, la composición atmosférica, la luz difusa que remite a Turner. La tierra, en cambio, ya no está en la materia del pigmento ni en la huella del trazo; se manifiesta como espesor perceptivo. En estas imágenes, el espectador siente una profundidad que no proviene del espacio representado, sino del desajuste entre la ilusión de lo natural y la evidencia de lo artificial. Allí se produce la pugna: el mundo abre su horizonte simbólico, pero la tierra, convertida en ilusión que se retrae, resiste, recordando que esa naturaleza es una proyección descontextualizada.

El concepto de espesor perceptivo permite describir esta experiencia sin recurrir a categorías materiales. En lugar de una tierra física, encontramos una tierra fenomenológica: una reserva de opacidad dentro de la transparencia digital. Como señala Donnarumma (2023), el arte algorítmico obtiene su fuerza estética en los puntos donde la representación falla, donde el sistema revela sus propios límites como parte de la imagen. En *Landscape Paintings*, esos límites aparecen en las transiciones imprecisas, en las zonas donde el algoritmo se aproxima demasiado a la perfección y genera una sensación de irrealidad. El espectador percibe esta falla no como defecto, sino como profundidad. Lo que resiste, lo que hace de suelo, es la imposibilidad de una naturaleza completamente representada. La verdad que salta al frente no es la del paisaje como copia del mundo, sino la del paisaje como signo de su pérdida.

Desde la filosofía del paisaje de Simmel, esta operación puede entenderse como una nueva forma de enmarcamiento (*Rahmung*). El algoritmo actúa como mirada cultural que selecciona fragmentos del mundo y los organiza en figura, pero el marco ya no delimita físicamente un campo de visión, sino que produce su propio horizonte virtual. Según Harrington (2004), el paisaje moderno expresaba la capacidad de la cultura para convertir lo indeterminado en totalidad. En Quayola, esa capacidad se automatiza: la cultura se inscribe en el código, y el código reinterpreta el mundo. Sin embargo, el efecto estético permanece, porque la mirada humana sigue siendo

convocada a integrar lo múltiple en unidad. El espectador, al recorrer la superficie pixelada, experimenta un movimiento de reconocimiento y pérdida: el paisaje se abre como mundo, pero su espesor se percibe en la ausencia del gesto humano que lo habría producido. La máquina configura, pero el sentido emerge en la mirada que busca el rastro de lo humano.

Esta dialéctica entre presencia y ausencia vincula directamente a Quayola con la hermenéutica heideggeriana. Heidegger (2002) señala que la verdad no es una propiedad de los entes, sino un acontecer de desocultamiento (*aletheia*). En *Landscape Paintings*, el desocultamiento ocurre precisamente en la tensión entre el cálculo técnico y la experiencia sensible. Lo que se revela no es el paisaje en sí, sino la condición contemporánea de nuestra relación con la imagen: el hecho de que lo visible se ha vuelto producto de sistemas que exceden nuestra percepción. Como sugiere Parikka (2015), el arte digital se visualiza en el funcionamiento material de la cultura algorítmica, transformando la estética en una arqueología de la visión. La verdad que se pone en obra en Quayola no es la de la naturaleza, sino la de la mediación técnica como forma de mundo.

El paisaje algorítmico, al igual que la obra heideggeriana, no busca resolver la pugna entre apertura y resistencia, sino sostenerla. En este sentido, *Landscape Paintings* puede interpretarse como una actualización de la estructura de mundo y tierra en el ámbito digital. El mundo es el horizonte simbólico que la obra hereda de la historia del arte; la tierra es la resistencia perceptiva que emerge del proceso técnico. La verdad acontece en la fricción entre ambos: en el reconocimiento simultáneo de lo cultural y lo inhumano, de lo sensible y lo calculado. Esta experiencia redefine la relación entre arte y técnica. La tecnología, que para Heidegger (1977) era una forma de desocultamiento instrumental, en Quayola se convierte en medio poético: un modo de revelar la tensión entre memoria y simulacro.

El espesor perceptivo se consolida aquí como una categoría estética que describe la condición contemporánea del paisaje digital. En el arte generado por inteligencia artificial, el espectador se enfrenta a imágenes que ya se saben mediadas, producidas por sistemas que asumen la multiplicidad

de lo posible. Existe una predisposición cultural a reconocer esta mediación como parte del acontecimiento estético: lo digital no oculta su ajenidad, sino que la convierte en superficie de contemplación. Sin embargo, en el motivo pictórico del paisaje ocurre algo distinto. Su autonomía histórica dentro del arte lo presenta como un lugar seguro, familiar, un territorio donde la mirada encuentra estabilidad incluso en medio del caos.

Cuando ese paisaje se descontextualiza y se desmaterializa en su forma algorítmica, se desestabiliza también el suelo sobre el que se apoya la experiencia estética: la tierra heideggeriana deja de ofrecer resistencia tangible. Lo que permanece es una sensación de densidad, un espesor perceptivo que opera como último punto de anclaje del espectador ante la inmaterialidad del mundo digital. Esa capa de profundidad no es ilusión óptica, sino conciencia de un nuevo tiempo histórico: un tiempo en el que el arte y la humanidad se reconocen a sí mismos como parte de un proceso continuo de mediación. En Quayola, esta experiencia no destruye la emoción estética, sino que la amplifica, porque la belleza del paisaje digital radica en esa fragilidad perceptiva donde lo real y lo tecnológico se confunden, dejando al espectador suspendido en la frontera entre memoria y cálculo.

Así, la importancia filosófica de Quayola radica, entonces, en que su trabajo no ilustra la teoría, sino que la prolonga. Sus paisajes no son experimentos formales, sino meditaciones visuales sobre la naturaleza de la verdad en el arte contemporáneo. Al reemplazar el trazo por el cálculo, no elimina la presencia de la tierra, sino que la traslada al terreno perceptivo: la tierra se vuelve extrañamiento en la experiencia, el lugar donde la visión tropieza con lo que no puede comprender. En ese tropiezo se juega el sentido del arte en la era de la inteligencia artificial. El paisaje, liberado del pigmento y del soporte, sigue siendo el escenario donde mundo y tierra se enfrentan.

Conclusiones

El recorrido heideggeriano y simmeliano permite comprender que el paisaje de IA conserva la estructura esencial de la experiencia de la filosofía del arte: apertura de mundo y resistencia de un fondo que preserva. La diferencia no reside en la desaparición de la pugna, sino en su desplazamiento. Allí donde

la tierra se hacía sentir como materialidad, ahora comparece como espesor perceptivo: una densidad fenomenológica que surge de la fricción entre memoria cultural y cálculo técnico. Esta reubicación no empobrece la verdad del arte; le otorga un nuevo régimen de aparición. En *Landscape Paintings*, Quayola encarna esa mutación al convocar la tradición paisajística y exponer, a la vez, su desanclaje algorítmico. Lo que acontece es una verdad de la mediación: el arte funda experiencia cuando sostiene la tensión entre lo que se abre y lo que se retira, incluso sin huella material.

De esta manera, proponer el espesor perceptivo como categoría estética contemporánea permite pensar la continuidad del paisaje más allá del soporte. El paisaje de IA no se limita a simular naturaleza; instala un campo hermenéutico donde la experiencia estética emerge del intervalo entre reconocimiento y pérdida. En ese intervalo se juega una promesa para la teoría del arte: comprender la técnica como condición de verdad y no solo como instrumento de representación.

Asimismo, reconocer el espesor perceptivo como desplazamiento de la tierra heideggeriana abre una vía para comprender cómo el arte digital reconfigura la relación entre lo visible y su trasfondo. La imagen algorítmica, al desmaterializar el soporte, no elimina la resistencia, sino que la redistribuye: lo que antes se alojaba en el pigmento o el trazo ahora se aloja en la conciencia de una distancia, en la insinuación de una ausencia que estructura la percepción. Esta dimensión fenomenológica demuestra que la experiencia estética contemporánea exige un nuevo vocabulario crítico capaz de captar la especificidad de la mediación técnica. En este sentido, el paisaje de IA no inaugura una ruptura con la tradición, sino que prolonga su pregunta fundamental: cómo permitir que el mundo se abra sin suprimir aquello que lo sostiene. Comprender esa tensión redefine la tarea de la estética en la era digital y sitúa al espesor perceptivo como una categoría clave para pensar el arte en un tiempo dominado por el cálculo.

Referencias

- Colapinto, A. (2005). Reliability, earth, and world in Heidegger's "The origin of the work of art." *Philosophy Today*, 49(Suppl.), 161–165. <https://doi.org/10.5840/philtoday200549Supplement20>
- Donnarumma, M. (2023). *Against the norm: Aesthetic inquiry into digital bodies*. MIT Press. https://marcodonnarumma.com/publications/Donnarumma_Against-the-Norm_JDCS23_author-personal-copy.pdf
- Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.
- Frisby, D. (1992). *Simmel and since: Essays on Georg Simmel's social theory*. Routledge.
- Hansen, M. B. N. (2020). *Feed-forward: On the future of twenty-first-century media*. University of Chicago Press.
- Harries, K. (2009). *Art matters: A critical commentary on Heidegger's "The origin of the work of art."* Springer.
- Harrington, A. (2004). *Art and social theory: Sociological arguments in aesthetics*. Polity Press.
- Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). Harper & Row.
- Heidegger, M. (2002). *Off the beaten track* (J. Young & K. Haynes, Eds. & Trans.). Cambridge University Press.
- Manovich, L. (2018). *AI aesthetics*. Strelka Press.
- Parikka, J. (2015). *A geology of media*. University of Minnesota Press.
- Simmel, G. (2013). *Filosofía del paisaje* (J. M. Álvarez, Trad.). Casimiro.
- Thomson, I. (2012). *Heidegger, art, and postmodernity*. Cambridge University Press.
- Young, J. (2001). *Heidegger's philosophy of art*. Cambridge University Press. <https://www.sfu.ca/~poitras/Heidegger-Philosophy-of-Art.pdf>

Inteligencia Artificial y su impacto en el trabajo cognitivo desde una perspectiva humanística

Artificial Intelligence and its impact on cognitive work from a humanistic perspective

María Aránzazu Serantes

arantxa.serantes@ufv.es

Universidad Francisco de Vitoria

ORCID ID 0000-0001-9403-7522

Resumen: En estos últimos años, la inteligencia artificial ha entrado en diferentes áreas de nuestra vida, incluyendo el ámbito educativo. Esta situación no solo presenta retos técnicos y pedagógicos, sino que cuestiona los fundamentos mismos de la educación: el propósito de enseñar, de adquirir conocimientos y, sobre todo, de razonar. La inteligencia artificial (IA) es un sistema que simula algunas capacidades del cerebro humano, como la resolución de problemas y la comprensión del lenguaje. Esto plantea una nueva forma de entender el trabajo mental como algo que moldea y enriquece nuestra humanidad.

En este artículo, se va a analizar cómo la inteligencia artificial afecta el trabajo mental, pero desde un enfoque más humano. Esto significa que vamos a considerar no solo lo que la IA puede hacer, sino también cómo influye en cómo aprendemos, en nuestra individualidad al aprender y en la importancia del profesor como guía en la transmisión cultural y educativa. También se considerará si es beneficioso estudiar y adquirir conocimientos con ayuda de estas tecnologías, teniendo en cuenta la importancia del pensamiento crítico, la creatividad y el desarrollo integral de la persona.

Palabras clave: IA; conocimiento; filosofía; diálogo; metacognición.

Abstract: In recent years, artificial intelligence has entered different areas of our lives, including education. This situation not only presents technical and pedagogical challenges but also questions the very foundations of education: the purpose of teaching, of acquiring knowledge and, above all, of reasoning. Artificial intelligence (AI) is a system that simulates some of the capabilities of the human brain, such as problem solving and language comprehension. This raises a new way of understanding mental work as something that shapes and enriches our humanity.

In this article, we will analyze how artificial intelligence affects mental work, but from a more human perspective. This means that we will consider not only what AI can do, but also how it influences how we learn, our individuality in learning, and the importance of the teacher as a guide in cultural and educational transmission. We will also consider whether it is beneficial to study and acquire knowledge with the help of these technologies, taking into account the importance of critical thinking, creativity, and the integral development of the person.

Keywords: AI; knowledge; philosophy; dialogue; metacognition.

La IA como herramienta cognitiva: ¿Asistencia o sustitución?

La introducción de la inteligencia artificial en entornos educativos ha generado opiniones encontradas. Algunos la ven como una posible amenaza para el pensamiento independiente, mientras que otros la consideran una oportunidad para mejorar la educación. Esta dicotomía nos lleva a reflexionar sobre el tipo de pensamiento que se fomenta en las aulas y el impacto de las herramientas digitales en su evolución.

Levy (1999) ya advertía que las tecnologías no son neutras y que, lejos de ser meros instrumentos, configuran modos de pensar y de conocer. Desde esta mirada, la IA no solo ofrece asistencia para resolver tareas, sino que moldea las formas de razonar, organizar ideas, jerarquizar información y plantear soluciones. Por ello, su integración debe ser crítica y consciente.

El debate actual sobre el trabajo cognitivo en contextos digitalizados ha evolucionado hacia enfoques que ven la mente como un sistema interconectado y estrechamente vinculado a objetos externos. A partir de la propuesta inicial de Clark y Chalmers (1998, 2018), la teoría de la mente extendida argumenta que las herramientas tecnológicas no solo aumentan los recursos para pensar, sino que se incorporan funcionalmente en los propios procesos de razonamiento, memoria y resolución de problemas. Este método ha sido retomado y profundizado por estudios recientes en alfabetización digital y automatización cognitiva (Selwyn, 2019; Williamson & Piattoeva, 2023), que examinan de qué manera la externalización sistemática de actividades intelectuales altera la esencia del esfuerzo cognitivo, trasladando ciertos procesos —como la búsqueda, la organización y la síntesis de información— hacia sistemas automatizados.

Del mismo modo, investigaciones sobre atención profunda, pensamiento crítico y economía cognitiva (Kahneman, 2011; Wu, 2020) señalan que la accesibilidad continua de sistemas inteligentes puede fomentar tipos de “cognición delegada”, donde el individuo disminuye su participación reflexiva a favor de respuestas producidas automáticamente. Desde una perspectiva humanista, estos procesos requieren reconsiderar la conexión entre autonomía, agencia y elaboración de conocimiento, ya que no toda expansión tecnológica se traduce en una mejora de la actividad intelectual. Así, la integración pedagógica de la IA debe tener en cuenta tanto su capacidad para crear andamiajes cognitivos así como su influencia en la creación de hábitos mentales, formación de criterios y disposiciones afectivo-reflexivas que respalden el pensamiento crítico.

El trabajo cognitivo —es decir, el esfuerzo mental destinado a comprender, razonar y crear— no se reduce a procesar información. Incluye también la capacidad de formular preguntas relevantes, encontrar relaciones entre elementos dispersos e imaginar alternativas. La IA puede ofrecer respuestas, pero no reemplaza el valor formativo de formular las preguntas. Aquí es donde la pedagogía humanista cobra protagonismo: no se trata de preparar autómatas eficientes, sino de educar personas capaces de comprender el mundo y transformarlo con responsabilidad.

Pensamiento lógico-creativo y el desafío de enseñar a pensar

El pensamiento lógico-creativo es una destreza esencial para convivir en sociedad. Es una forma de inteligencia que combina la lógica estructurada con la habilidad de crear ideas novedosas y que posibilita solucionar problemas de un modo pertinente, flexible y original (Guilford, 1950; Piaget, 1970). No se puede limitar la enseñanza del pensamiento a seguir algoritmos o copiar esquemas mentales; más bien, requiere que el alumno esté en el centro del proceso, animándolo a investigar, experimentar, cometer errores y reelaborar.

En este sentido, la IA tiene el potencial de ser un agente relevante para promover el pensamiento creativo, sobre todo cuando funciona como interlocutora flexible que puede brindar retroalimentación, sugerir retos o ampliar las posibilidades de acción del estudiante. Sin embargo, su utilización no garantiza automáticamente un aprendizaje significativo. Como muestran algunas experiencias de aula, si no se acompaña el uso de IA con orientación docente y un propósito pedagógico claro, existe el riesgo de que los estudiantes se conviertan en consumidores pasivos de soluciones automáticas.

El uso humanístico de la inteligencia artificial en la educación debe enfocarse en aumentar la independencia intelectual del alumno. En otras palabras, ayudarle a construir su propio entendimiento, tomar decisiones basadas en argumentos y aumentar su autorregulación. Según Paulo Freire (1970), educar es un acto de libertad, no de domesticación. Siguiendo este razonamiento, la inteligencia artificial debe ser una aliada de la emancipación cognitiva, no un medio para sustituirla.

¿Vale la pena construir conocimiento asistidos por IA?

Para responder de manera afirmativa a esta pregunta, es necesario matizar el concepto de "asistencia". Si se considera como una guía que respalda, incentiva y complementa el proceso cognitivo del alumno, entonces sí: la IA puede ser un instrumento poderoso. Sin embargo, si su utilización se restringe a resolver ejercicios, producir textos o automatizar respuestas sin que haya

reflexión o asimilación de conceptos, entonces se vuelve un medio de empobrecimiento intelectual.

Según Litwin (2005), el saber, desde el punto de vista del humanismo, no es simplemente una acumulación de datos, sino que se trata de una construcción activa, contextualizada, mediada por la sociedad y con un enfoque ético. En este contexto, la IA tiene la capacidad de permitir el acceso a información, plantear situaciones simuladas, crear actividades personalizadas o desempeñar el papel de un tutor virtual. No obstante, el peligro está en que los alumnos confundan la elaboración de respuestas con la formación del conocimiento.

Es en este punto cuando la función del docente cobra una nueva dimensión. En lugar de ser simplemente el transmisor de contenidos, se transforma en un curador pedagógico, un mediador cultural y un orientador ético. Ya no se limita a enseñar "lo que Google no ofrece", sino que tiene la responsabilidad de acompañar procesos de formación en los que los alumnos aprenden a pensar en conjunto con otros, a distinguir lo verdadero de lo falso, lo pertinente de lo superfluo y lo justo de lo injusto. En este contexto, merece la pena estudiar con ayuda de inteligencia artificial, siempre que el fin no sea solo la eficiencia, sino también la educación integral del individuo.

Emociones, aprendizaje y la dimensión afectiva de la cognición

La incorporación del aspecto emocional en el proceso de aprendizaje es una de las contribuciones más significativas del enfoque humanista. Según varias investigaciones en neuroeducación, las emociones no son un impedimento para la razón; por el contrario, la orientan y condicionan (Bueno, 2018; Gago Galvagno & Elgier, 2018). Sentir es un elemento esencial de la experiencia cognitiva y aprender implica sentir.

En este contexto, la implementación de la IA en el aula brinda nuevas oportunidades para abordar los elementos emocionales del aprendizaje, particularmente si los sistemas están creados para ajustarse al tono comunicativo del alumno, proporcionar asistencia y manejar la frustración, como si de un *coach* se tratara. No obstante, la empatía simulada que producen

ciertos sistemas no debe ser confundida con una auténtica relación humana. La inteligencia artificial tiene la capacidad de imitar emociones, pero no de sentirlas.

Resulta esencial, en este caso, diferenciar entre el vínculo educativo y el acompañamiento emocional. Cuando el alumno muestra frustración, la inteligencia artificial puede dar respuestas de aliento; sin embargo, no es capaz de sustituir la escucha activa, la mirada comprensiva o la empatía de un profesor comprometido. Por lo tanto, la inteligencia artificial debe ser concebida como un complemento que favorece el bienestar emocional del alumno, pero no como un reemplazo de la relación pedagógica.

Ética, subjetividad y pensamiento crítico: una alerta necesaria

El efecto de la IA en la subjetividad de los alumnos no debe minimizarse. Así como los entornos digitales han cambiado la forma en que nos comunicamos, nos informamos o consumimos, también están alterando nuestra manera de pensar. En este contexto, educar en la época de la inteligencia artificial supone un exigente deber ético: instruir a individuos críticos que sean capaces de cuestionar lo que se dice, plantearse problemas sobre el mundo y resistirse a la lógica del pensamiento único.

La disminución de la autonomía intelectual es el peligro de delegar las tareas cognitivas a sistemas automatizados. Nicholas Carr (2010) advierte que, si nos habituamos a que las máquinas piensen por nosotros, dejamos de ser capaces de concentrarnos, hacer un análisis profundo y emitir un juicio personal. En el contexto educativo, esto puede manifestarse en una cultura de atajos, de obtener resultados inmediatos sin entender el proceso.

Por lo tanto, se vuelve necesario que se implemente una alfabetización crítica en IA. No es suficiente con instruir sobre cómo utilizar estas herramientas; también se necesita enseñar a cuestionarlas. ¿Con qué información trabajan? ¿Cuáles son los sesgos que tienen? ¿Qué intereses simbolizan? ¿Qué efecto tienen nuestras decisiones? Estas preguntas tienen que ser parte de una ética renovada en términos del conocimiento, donde el

pensamiento crítico no sea solo otra competencia, sino el núcleo de la educación.

La IA en el aula: una oportunidad si se media con sentido pedagógico

La integración de la inteligencia artificial en el aula no debería ser vista como una mera actualización tecnológica, sino como una transformación del ecosistema educativo. Si se diseña con un enfoque accesible, la inteligencia artificial (IA) puede favorecer la inclusión educativa, adaptarse al ritmo y estilo de cada alumno, proporcionar retroalimentación instantánea y posibilitar maneras innovadoras de personalizar el aprendizaje.

No obstante, como correctamente indica Litwin (2005), no toda integración tecnológica en el aula tiene un impacto significativo. Para que sea así, debe tener un objetivo educativo definido, estar dentro de un plan didáctico coherente y ser evaluada en términos de sus impactos formativos. Si no es así, existe el peligro de que se incurra en un "tecnologismo vacío", en el que la innovación desplaza a la pedagogía.

Un enfoque humanístico de la IA en educación pone en el centro a la persona que aprende. No se trata de educar para el mercado, sino para la vida; no de preparar usuarios eficaces, sino ciudadanos que sean críticos, empáticos y responsables. En esta línea, el trabajo cognitivo respaldado por IA debe integrarse en un proyecto educativo más extenso, en el que la tecnología sirva a la dignidad del ser humano y no al contrario.

Posibles escenarios desde un punto de vista cognitivo bajo una perspectiva humanista

Escenario 1: Escritura académica asistida con IA

En el campo de la producción escrita, el apoyo de la IA se ha transformado en una herramienta comúnmente empleada para crear esquemas, recomendar enfoques o presentar estructuras argumentativas. Tomemos el ejemplo de una estudiante de filosofía que desarrolla un ensayo sobre la libertad conforme a la corriente existencialista y utiliza un modelo generativo para crear un mapa conceptual preliminar.

Desde la óptica del trabajo cognitivo, la implementación de IA proporciona una externalización operativa de funciones preparatorias — como la estructuración inicial de ideas o la detección de relaciones conceptuales—, lo que permite que el esfuerzo mental se dirija a actividades de nivel superior: cuestionar, reinterpretar o establecer vínculos críticos entre fuentes primarias. No obstante, esta externalización únicamente tiene carácter formativo si la estudiante retiene el control epistemológico del proceso: elige, compara, ajusta y reconfigura lo que la herramienta sugiere.

En cuanto a autonomía, la mediación docente desempeña un papel fundamental para impedir que la IA reemplace las habilidades de análisis o argumentación. La estudiante debe tomar la responsabilidad de fundamentar cada decisión conceptual, valorar la relevancia de las propuestas creadas y aceptar las limitaciones propias del contenido generado automáticamente.

Desde una perspectiva ética y humanista, este contexto demuestra que la IA no sustituye la interpretación filosófica, sino que sirve como un apoyo que amplía el horizonte de oportunidades, siempre que esté enfocada en el objetivo formativo de fomentar la agencia cognitiva. El aprendizaje significativo surge, de este modo, de la interacción entre la herramienta y la reflexión activa del individuo y no de la simple repetición de respuestas automáticas.

Escenario 2: Resolución de dilemas éticos generados por IA

En un curso de ética, el profesor emplea IA para crear dilemas que faciliten la aplicación de teorías morales aprendidas en clase. El sistema presenta un caso que abarca situaciones relacionadas con la privacidad, seguridad y autonomía en un entorno de supervisión algorítmica.

En este contexto, la IA actúa como creadora de incentivos cognitivos que aumentan la complejidad de la situación. Sin embargo, el entendimiento pleno del dilema no está en la herramienta, sino en la actividad reflexiva del alumnado: reconocer principios en conflicto, aplicar marcos teóricos, justificar posturas y evaluar consecuencias. En este contexto, la externalización se restringe a la creación del caso, mientras que la formulación

de argumentos y la adopción de una postura siguen siendo actividades principalmente humanas.

Este tipo de actividades refuerza la independencia intelectual, puesto que los estudiantes deben analizar de manera crítica las propuestas de la IA, identificar posibles prejuicios en la forma de presentar el dilema y reconfigurar el problema desde su propia visión. De este modo, la utilización de IA se transforma en un objeto de estudio y en una herramienta para investigar la relación entre ética y tecnología.

Desde la perspectiva humanista, la práctica muestra que la IA impulsa el pensamiento crítico solo si se combina con métodos pedagógicos que promueven el juicio, la reflexión y la comprensión contextual de las cuestiones éticas. La herramienta no guía el pensamiento: lo estimula, lo desafía y lo expande, siempre que el estudiante conserve su protagonismo epistemológico, como ocurre con herramientas como *Habla con la Historia* (<https://hablaconlahistoria.es/>) o *Flowhunt* (<https://www.flowhunt.io>) entre otros, que promueven el diálogo socrático.

Conclusiones

La inteligencia artificial en la educación sólo es beneficiosa bajo un enfoque humanista. La IA, por sí sola, no asegura procesos de entendimiento profundo ni el impulso de la autonomía; su influencia depende más bien del tipo de mediaciones educativas que la rodean y del papel que se le asigna en la estructura del aprendizaje.

Los escenarios expuestos indican que los sistemas inteligentes pueden ayudar en la exploración conceptual, la organización del pensamiento y la creación de situaciones problemáticas que favorezcan el análisis crítico. No obstante, estas ventajas solo surgen cuando los estudiantes ejercen el control interpretativo, defienden su criterio y realizan una deliberación responsable acerca de los contenidos generados automáticamente.

En resumen, sería posible afirmar que la IA puede potenciar la experiencia educativa así como el trabajo intelectual, siempre que se interprete

como una herramienta de apoyo que respeta la individualidad del sujeto y su capacidad de análisis. El enfoque no está en la efectividad técnica, sino en retomar una enseñanza basada en la pregunta, el diálogo y la formación ética. Por lo tanto, la labor educativa no debe limitarse a aceptar sin cuestionar los automatismos, sino a crear contextos de significado que posibiliten a las personas reflexionar con otros —sean humanos o entidades artificiales— sin perder su agencia, su creatividad ni su responsabilidad en el desarrollo del conocimiento.

Dar respuesta a todo ello, desde una perspectiva innovadora, supone examinar nuestras prácticas, reconsiderar el propósito de la enseñanza y confirmar la posición de la *paideia* como un espacio para el encuentro, diálogo y formación humana. La tecnología no resuelve por sí sola las dificultades educativas, como señala Litwin (2005); la diferencia radica en el enfoque pedagógico que se le asigna.

En definitiva, sí vale la pena construir conocimiento asistidos por la inteligencia artificial, pero solo si lo hacemos con una ética del cuidado, una pedagogía de la pregunta y una mirada puesta en la dignidad del sujeto que aprende.

Referencias

- Bueno, D. (2018, 24 de enero). *¿Cómo cambia nuestro cerebro al aprender?* [Vídeo]. YouTube. <https://www.youtube.com/watch?v=nXQe7I5WBXs>
- Carr, N. (2010). *The shallows: What the Internet is doing to our brains*. W. W. Norton & Company.
- Clark, A. (2018). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Dennett, D. (2017). *From bacteria to Bach and back: The evolution of minds*. W. W. Norton & Company.
- Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.
- Freire, P. (1970). *Pedagogía del oprimido*. Siglo XXI Editores.

- Gago Galvagno, L. G., & Elgier, Á. M. (2018). Trazando puentes entre las neurociencias y la educación. *Psicogente*, 21(40), 476–494. <https://doi.org/10.17081/psico.21.40.3087>
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. <https://doi.org/10.1037/h0063487>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Levy, P. (1999). *Cibercultura*. Paidós.
- Litwin, E. (2005, 30 de junio – 2 de julio). *De caminos, puentes y atajos: el lugar de la tecnología en la enseñanza* [Conferencia]. II Congreso Iberoamericano de EducaRed: “Educación y Nuevas Tecnologías”, Buenos Aires, Argentina. <https://www.fhcs.unp.edu.ar/practicaprofesional/wp-content/uploads/2019/03/Litwin-Edith-Texto-ampliatorio.pdf>
- Piaget, J. (1970). *La psicología de la inteligencia*. Morata.
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.
- Williamson, B. (2017). *Big data in education*. SAGE.
- Williamson, B., & Piattoeva, N. (2023). *Education governance and datafication: Critical perspectives*. Bloomsbury Academic.
- Wu, T. (2020). *The attention merchants: The epic scramble to get inside our heads*. Vintage Books.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

**El sistema global de la Inteligencia Artificial
Generativa.
Trabajo, poder y desafíos en la segunda oleada
informacional**
The global system of Generative Artificial Intelligence.
Work, power, and challenges in the second wave of
information

Patricio Feldman

patofeldman@gmail.com

Instituto de Investigaciones Gino Germani, Universidad de Buenos Aires

ORCID ID 0000-0001-7012-375X

Martín Gendler

martin.gendler@gmail.com

Instituto de Investigaciones Gino Germani, Universidad de Buenos Aires

ORCID ID 0000-0001-5781-6367

Resumen: El artículo busca profundizar en las características, problemáticas, actores y desafíos éticos y regulatorios de la Inteligencia Artificial Generativa (IAGen), uno de los desarrollos clave de la segunda oleada informacional, entendiéndola como una tecnología estructurada en un sistema de características globales. Se inicia definiendo a la IAGen en torno a sus características diferenciales para luego, por medio de una metodología cualitativa de análisis y revisión bibliográfica, proponer y desplegar una sistematización clasificatoria de siete capas (software, mano de obra, datos, hardware, infraestructura, materiales y energía) interseccionadas con los niveles macro, meso y microsociales. Esto permite ahondar tanto en las características como también en los desarrollos tecnológicos, recursos naturales, países, empresas y otros actores implicados de cada capa, constituyendo una base analítica firme para focalizar en las

problemáticas y desafíos éticos y regulatorios, comunes y diferenciales, que se presentan en cada una de las mismas. Su relevancia se encuentra en el desarrollo de una herramienta teórico-conceptual para el análisis de los desafíos, problemáticas, oportunidades y características que presenta una de las principales tecnologías de esta nueva etapa informacional.

Palabras clave: Inteligencia Artificial Generativa; internet; tecnologías digitales; paradigma 4.0.

Abstract: This article seeks to delve into the characteristics, issues, actors, and ethical and regulatory challenges of Generative Artificial Intelligence (GenAI), one of the key developments of the second information wave, understanding it as a technology structured in a system of global characteristics. It begins by defining the GenAI around its differential characteristics and then, through a qualitative methodology of analysis and bibliographic review, proposes and deploys a classificatory systematization of seven layers (software, labor, data, hardware, infrastructure, materials, and energy) intersected with macro, meso and micro-social levels. This allows to delve deeper into the characteristics as well as the technological developments, natural resources, countries, companies, and other actors involved in each layer, constituting a solid analytical basis for focusing on the ethical and regulatory issues and challenges that arise in each of them. Its relevance lies in the development of a theoretical-conceptual instrument for the analysis of the challenges, problems, opportunities, and characteristics presented by one of the main technologies of this new informational stage.

Keywords: Generative Artificial Intelligence; internet; digital technologies; paradigm 4.0.

Introducción

Los avances recientes de la Inteligencia Artificial Generativa (IAGen), un desarrollo de la Inteligencia Artificial (IA) asociada al *Meaching Learning*, *Deep Learning* y Procesamiento de Lenguaje Natural, entre otras, una de las tecnologías más relevantes del paradigma 4.0¹ (Girolimo & Feldman, 2021), introducen debates de todo tipo que incluyen una discusión ontológica en torno al carácter particular de una nueva agencia, los debates en torno al futuro del trabajo en el marco de la automatización de tareas cognitivas/intelectuales y otros aspectos como la geopolítica de la IA asociada a la disputa por la hegemonía global entre Estados Unidos y China. Este artículo recupera los enfoques de Crawford (2023) y Estampa (2024) que reconocen la necesidad de alinear el análisis en la economía política de la IAGen. En este sentido, se plantea como necesario visibilizar cuáles son las capas involucradas en el desarrollo de esta tecnología, sus expresiones territoriales específicas y el modo en que distribuyen las externalidades ocasionadas por el despliegue de un sistema global que ordena y estructura el trabajo, el poder y la generación de riqueza en la etapa actual del capitalismo informacional.

Partiendo de este último enfoque, se plantea que la IAGen es mucho más que solo los softwares que se utilizan comúnmente, dado que, en torno ella se estructura un sistema global que contiene siete capas, donde cada una puede ser analizada como nodo de una red mundial. Si bien, los nodos presentan cierta autonomía- una red de actores sociales propia, criterios organizativos particulares y una trayectoria histórica puntual- no existe

¹ El paradigma 4.0, también considerado como una segunda “fase” o “etapa” del informacionalismo (Zukerfeld, 2020), representa la profundización y complejización del paradigma informacional inaugurado en la década de 1970, pudiendo situarse su año de inicio en 2008 tras la crisis económica mundial, los juicios de propiedad intelectual y, muy especialmente, la masificación del smartphone y del ecosistema de plataformas digitales (Gendler, 2024a). Cabe destacar que los saltos de escala de los procesos de datificación social que conlleva esta nueva fase se conjugan con una batería de tecnologías digitales vistas como “clave” de la etapa: la inteligencia artificial, la internet de las cosas, la impresión 3D, la realidad aumentada, realidad virtual, entre muchas otras (Schwab, 2016).

posibilidad de que la red funcione sin una articulación fluida, estable y sólida entre cada uno de ellos. Estas capas organizan los mercados, el trabajo y la base material de una nueva oleada de desarrollo informacional. Se trata de un nuevo ensamblaje de bits y átomos de escala planetaria que abre una serie de tensiones, desafíos, amenazas y oportunidades, presentes y futuras, que van redefiniendo la relación entre tecnología y sociedad. La comprensión de este sistema global se considera de vital importancia para poder complejizar no solo el entendimiento de la IAGen, sino también lo que respecta a su análisis, implementación y posible regulación. En este sentido, el poder ahondar en el entendimiento no solo de las características y problemáticas del software central de la IAGen, los LLM, sino también contemplar los demás niveles, componentes, actores y ubicaciones geográficas que lo constituyen y sostienen, se considera central para una comprensión integral que pueda aportar mayor complejidad al campo de estudios.

Asimismo, las siete capas se cruzan con tres niveles organizativos de la sociedad: macro, meso y micro. Una particularidad de este sistema global, como plantea Costa et al. (2023), es que el nivel macro no corresponde a los Estados ni a los ámbitos internacionales de coordinación entre estos, sino que es ocupado por las grandes empresas de tecnología. Tanto el complejo GAFAM (Google, Amazon, Facebook, Apple, y Microsoft) y sus subsidiarias como OpenAI, para el caso norteamericano, como BATH (Baidu, Alibaba, Tencent, y Huawei), en el caso chino, desempeñan un protagonismo indiscutible en este nivel. Luego, en un nivel meso, se destacan los Estados Nacionales en sus diferentes niveles territoriales y una variedad de organismos internacionales, la mayoría dependiente de Naciones Unidas. Finalmente, a nivel micro, aparecen las organizaciones de la sociedad civil, asociaciones empresarias, pequeñas y medianas empresas y los ciudadanos. Precisamente por el tipo de entrecruzamiento de las siete capas con los tres niveles, la gobernanza de este sistema global se torna dificultosa, desordenada y fundamentalmente asimétrica.

El objetivo del presente trabajo es ahondar en el entendimiento de las características, problemáticas, actores y desafíos políticos y regulatorios de la IAGen a nivel global, considerándola una de las principales tecnologías

digitales disruptivas del paradigma 4.0. Para ello, se inicia el escrito definiendo la IAGen en torno a sus características particulares y dimensiones y su vínculo con otros desarrollos de IA relacionados. Tras esto, empleando una metodología cualitativa de análisis y revisión bibliográfica (Valles, 1999), se recurre al armado y despliegue de una sistematización transversal a los niveles macro, meso y microsociales, que permita clasificar a la IAGen en siete capas, considerando sus distintos niveles de incumbencia, las tecnologías y recursos sociotécnicos y naturales involucrados, los principales países implicados en cada capa, las empresas protagonistas, otros actores.

Su relevancia consiste en disponer de una herramienta teórico-conceptual pero también analítica útil para el estudio y entendimiento integral de las características, problemáticas y desafíos que protagonizan la actual etapa del capitalismo informacional.

La IA Generativa: la Nueva Joya del Informacionalismo

La IA Generativa (IAGen) constituye un desarrollo que combina avances previos² dentro del subcampo del *deep learning*, el cual, a su vez, forma parte del *machine learning*³. Se puede caracterizar como una tecnología de propósito general, esto es: realiza una función genérica que resulta fundamental para un gran número de productos o sistemas de producción, presenta rasgos de un fuerte dinamismo tecnológico, va mejorando a lo largo del tiempo, impulsando innovaciones sobre esa función genérica, y, además, resulta

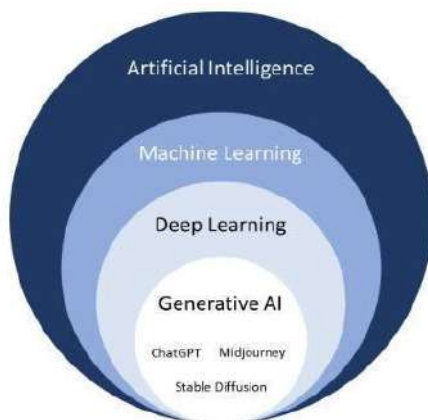
² Entre ellos: las redes neuronales profundas, las redes generativas antagónicas, el procesamiento de lenguaje natural, el aprendizaje por transferencia; en conjunto con otros desarrollos provenientes del machine learning como la optimización algorítmica, el modelado estadístico y el etiquetado de datos, entre otros.

³ El machine learning es una rama de la IA que permite a las máquinas aprender patrones a partir de datos y mejorar su desempeño sin ser programadas explícitamente. Incluye una variedad de algoritmos (como árboles de decisión o SVM) y abarca distintos enfoques (supervisados, no supervisados y por refuerzo). Por su lado, el deep learning es un subcampo del machine learning que utiliza redes neuronales artificiales con múltiples capas, lo que le permite aprender patrones más complejos y abstractos. Asimismo, por su alto nivel de complejidad, permite alcanzar mayores niveles de profundidad en el análisis de relaciones, características y predicciones (Madakam et al., 2022).

complementaria con otras innovaciones, lo cual facilita que se disemine en toda la economía y sociedad (Rosenberg & Trajtenberg, 2004).

Figura 1

(algunas) ramas, subcampos y desarrollos de Inteligencia Artificial



Nota. Yuan (2022).

La IAGen presenta cinco características que la distinguen de los otros campos y que permiten sostener que se trata de una IA con mayor potencial disruptivo que sus antecesoras. En primer lugar, no se trata de una IA acotada a una función específica, sino que permite realizar múltiples funciones y responder a diversos requerimientos que van desde la búsqueda de información hasta la generación de texto, análisis de datos, generación de imágenes, videos, entre otras opciones.

Un segundo componente refiere a la capacidad de generar contenidos (textos, programas, códigos, imágenes, videos, entre otros). De allí el término *generativa*. La forma en la que genera ese contenido es a través de instrucciones (*prompt*), en diálogo con los datos que han sido utilizados para entrenar los modelos, lo que demuestra que, como toda tecnología digital, incluye un tipo de interfaz humano-máquina-software.

La tercera característica se asocia al comportamiento extractivo. La IAGen necesita de manera constante y a una escala cada vez mayor, expandir la frontera de extracción, tanto sobre recursos naturales como datos, información y conocimiento.

Una cuarta dimensión refiere a la estructura de producción de conocimiento e innovación dentro de la IAGen. Se trata de un sistema basado en el *open source*, pero con una lógica de acaparamiento de la información y el conocimiento, tanto el producido en las instituciones de ciencia y tecnología, como también los datos generados en la web y en las plataformas digitales. La recolección de estos datos se orienta al perfeccionamiento de los algoritmos, sobre el cual, luego, se monta un modelo de negocios. Tanto estos algoritmos, como varias de las bases de datos utilizadas, son propiedad de pocas empresas, y su desarrollo no es compartido con otros agentes, y mucho menos aún con los propios usuarios.

Una última característica se vincula con la condición de tecnología de alto riesgo. Como indican Costa et al. (2023) y Vercelli (2024), considerar a la IAGen como de alto riesgo, supone incidir en todas las etapas del ciclo de vida. En la siguiente Tabla (Nº1), se pueden observar las principales características de la IAGen:

Tabla 1

Características de la IAGen

Propósito general	Abocada a realizar múltiples funciones y operaciones (generación de contenido, análisis de datos, programación, síntesis, operaciones de cálculo, entre otros).
Generativa	Destinada a generar contenido en múltiples formatos: texto, imagen, video, código, entre otros.
Extractiva	El contenido que generan y las operaciones que realizan parten de la necesidad de contar con una base de datos lo suficientemente grande para entrenar a los algoritmos. A mayor caudal y calidad de datos, mejor capacidad operativa de los algoritmos. Además de los datos, las empresas que producen estos algoritmos extraen conocimiento disponible en la red. Por último, para almacenar y procesar los datos y que las aplicaciones sean utilizadas por la población, se requiere extender la frontera de extracción sobre los recursos naturales.
Cerrada	Si bien usufructúan los entornos de código abierto, los algoritmos y los procesos centrales de la IAGen —en sus diferentes fases de desarrollo— están sujetos a derechos de propiedad intelectual y permanecen ocultos e inaccesibles para usuarios y organismos de control.

Alto riesgo	Tanto por sus impactos sociales, éticos, económicos y ambientales, como por las consecuencias que puede generar su uso respecto a diferentes dimensiones de la vida humana, se puede considerar a esta tecnología como de alto riesgo.
--------------------	--

Nota. Elaboración propia en base a Costa et al. (2023) y Crawford (2023).

El sistema global de la IAGen

El desarrollo de la IAGen se suele analizar desde la óptica de las empresas innovadoras, radicadas mayoritariamente en las economías centrales, abocadas al desarrollo de algoritmos con un poder de cómputo cada vez mayor, y por el otro lado, desde el punto de vista de los usuarios repartidos por el mundo que adoptan cada vez más esta tecnología. Asimismo, paulatinamente van surgiendo análisis y estudios enfocados en los efectos de la IAGen en distintas esferas, como la educación, la salud, el trabajo, la gestión urbana, entre otros. Sin embargo, aún son escasos los enfoques que se propongan abordar a la IAGen como una tecnología de producción global, sobre la cual se construye una red de actores, estatales y no estatales, del sector público y privado, humanos y objetos técnicos digitales y analógicos (Gendler, 2023a), distribuidos a lo largo y ancho del mundo, de cuya interacción depende el desarrollo de las variantes, tipos y aplicaciones de IAGen disponibles. Esto implica que los algoritmos, las bases de datos y otros elementos centrales no se pueden disociar del resto de los componentes que permiten que la IAGen funcione, se perfeccione, y pueda ser utilizada por individuos, empresas y gobiernos.

Siguiendo a Estampa (2024), a Crawford y Loger (2018), y agregando varios añadidos propios, se puede desagregar los componentes del sistema global de IAGen en siete capas. Entendemos por capa, distintos elementos que conforman el sistema global de IAGen, que se pueden analizar de manera individual, pero su función dentro del sistema está definida por la interdependencia e interconexión con las otras. Esto quiere decir que cualquier falla o ausencia en alguno de los nodos afecta al conjunto del sistema. Debido a la complejidad sociotécnica de sus elementos componentes, cada capa involucra dimensiones técnicas, económicas, políticas, culturales, institucionales y/o normativas, si bien, en algunos casos, alguna puede presentar mayor presencia que otra. Asimismo, cada una

presenta una expresión territorial específica que hace alusión a los territorios predominantes (regiones, países, ciudades) integrados a cada capa del sistema. Si bien, el sistema global de la IAGen adopta por definición un principio de expansión global, lo que dificulta asociar en algunas capas actividades con territorios, se puede identificar la predominancia territorial en la mayoría de los nodos de la red. A continuación (Tabla N°2), se enumeran las capas que componen el sistema global de la IAGen:

Tabla 2

Capas del sistema global de la IAGen

Capa	Título	Descripción	Tecnologías involucradas	(Principales) países involucrados	(Principales) empresas involucradas	Otros actores involucrados
1	Software	Algoritmos para procesamiento de lenguaje natural, visión artificial, recomendación algorítmica, producción generativa, etc.	Aprendizaje automático, redes neuronales, procesamiento de lenguaje natural, LLM	Estados Unidos, China	Google, Microsoft, OpenAI, Meta, Baidu, Alibaba, Tencent	Estados (a través de fondos de investigación), universidades, startups
2	Mano de obra	Microtarefas de etiquetado, <i>scoring</i> y moderación de contenido	Crowdsourcing	India, Líbano, Kenia, Uganda	Amazon Mechanical Turk, Upwork, TaskRabbit	Plataformas de <i>crowdsourcing</i> , empresas de subcontratación, trabajadores “independientes”
3	Datos	Datos de internet, biométricos y de salud, recolectados por OTD 4.0 (IoT, sensores y otros)	Big data, minería de datos, sistema R.A.P.P.A	Global	Google, Facebook, Amazon, Alibaba, Tencent (y otras)	Estados, agencias de inteligencia, organizaciones de derechos civiles

Cap a	Título	Descripción	Tecnologías involucradas	(Principal es) países involucra dos	(Principales) empresas involucrada s	Otros actores involucrados
4	Hardware	Dispositivos, semiconductores y componentes electrónicos	Microprocesadores, GPUs, memorias	Taiwán, China, Corea del Sur, Estados Unidos	Intel, NVIDIA, TSMC, Samsung	Estados (a través de subsidios y regulaciones), consorcios de investigación, organismos de estándares
5	Infraestructura	Centros de datos, redes de conectividad	Computación en la nube, 5G, fibra óptica	Estados Unidos, China, Noruega, Suecia	Amazon Web Services, Microsoft Azure, Google Cloud, Nokia, Huawei	Estados (a través de la regulación de telecomunicaciones), proveedores de servicios de internet (ISP), UIT, 3GPP
6	Materiales	Minerales y materiales raros para componentes electrónicos	Litio, cobalto, tierras raras	República Democrática del Congo, Brasil, Perú	Empresas mineras multinacionales, fabricantes de baterías	Estados (a través de políticas mineras), comunidades locales
7	Energía	Fuentes de energía renovables y no renovables	Energía solar, eólica, nuclear, gas, petróleo	Global	Empresas energéticas, proveedores de servicios públicos	Estados (a través de políticas energéticas), organizaciones ambientalistas

Nota. Elaboración propia en base a Estampa (2024); y Feldman (2024).

La primera capa corresponde al *software*⁴. Esos desarrollos son creados y desplegados por las Big Tech y una red de *startups* que circulan bajo su órbita, predominando las empresas norteamericanas y chinas. Más allá de que el sector privado lidera estos desarrollos, los Estados de ambas potencias despliegan una variedad de instrumentos que potencian ese crecimiento: exenciones impositivas, formación de investigadores, científicos, y programadores, protección comercial, entre otras. Un debate central radica en la posibilidad de regular estos desarrollos durante todo el ciclo de vida. La principal complicación es que las regulaciones las ejercen los Estados, que

⁴ Comprende, a los LLM (ChatGPT, Claude, Gemini, DeepSeek, entre otros ejemplos) y a los principales desarrollos de IA que los hacen posibles mencionados previamente.

ocupan el nivel meso, mientras que las empresas a ser reguladas, dominan el nivel macro. Emerge, por lo tanto, una tensión entre capacidades estatales, niveles organizativos, responsabilidades, y el establecimiento de normas que permitan encuadrar esta capa.

En segundo lugar, se destaca la *mano de obra*. Incluye una serie de micro tareas asociadas al refinamiento de los algoritmos y/o la moderación de contenidos, mayormente tercerizadas a través de redes de *crowdsourcing*⁵ (Crawford, 2023). La mano de obra utilizada para realizar estas tareas se concentra en diferentes países como India, Líbano, Kenia y Uganda, siendo contratada en otras latitudes. Se genera una relación triangular entre cliente, plataforma, y trabajadores, en la cual, estos últimos no son reconocidos como tal, sino bajo la figura de “empleados autónomos”, sujetos a un proceso de gestión algorítmica centralizada.

La tercera capa constituye la materia prima digital principal de todo el sistema global (Srnicek, 2018): los *datos* y sus procesos de extracción, correlación, y procesamiento. Estos procesos se desarrollan a escala global, principalmente en la web y las plataformas digitales, mayormente ejercido por los dueños de estas últimas: las GAFAM, las BATH y otras como Netflix, Spotify, etc. Es posible considerar que todo el ecosistema de tecnologías digitales se ha ido adaptando para eficientizar los procesos, evidenciándose una mayor extracción en regiones, ciudades, y países con regulaciones laxas que permiten que empresas operen con mayor facilidad, o bien en contextos donde es facilitada o efectuada directamente por los gobiernos. En este caso, los niveles macro y meso se entremezclan.

⁵ La plataforma más emblemática en este aspecto es *Amazon Mechanical Turk*, aunque existen otras como *Clickworker* y *Microworker*. Los trabajadores acceden a pequeñas tareas que le son asignadas en función de la cantidad de tareas aceptadas y el puntaje obtenido en cada una de ellas. Los honorarios dependen del cliente y el tipo de tarea, pero no de la reputación adquirida por los trabajadores (Reynolds et al., 2024).

En cuarto lugar, se destaca el *hardware*⁶. En este caso, tres países lideran ampliamente este segmento: Taiwán, China, y Corea del Sur. A pesar de un incipiente proceso de *reshoring*, las principales empresas norteamericanas (Apple, por ejemplo) aún producen sus dispositivos en el gigante asiático. La empresa más importante en este rubro es Foxconn, de origen taiwanés, pero que centra sus operaciones en territorio chino. En Corea del Sur, Samsung y SK Hynix lideran la producción de chips. Sin embargo, más allá de donde se elaboran estos componentes, una de las empresas más importantes de fabricación de semiconductores y GPU, lo que le otorga el verdadero poder de cómputo a la infraestructura y los dispositivos digitales, es NVIDIA, multinacional norteamericana.

La quinta capa corresponde a la *infraestructura*. Los centros de datos más grandes a nivel mundial se distribuyen en Estados Unidos, China y Europa. Los servicios de computación en la nube son provistos en el mundo occidental por tres empresas: Amazon Web Services, Google Cloud, y Microsoft Azure, destacándose en China, Alibaba, Huawei y Tencent Cloud (Hawkins et al, 2025). Son estas corporaciones quienes controlan el grueso de la infraestructura que sostiene el crecimiento y desarrollo de la IAGen. Más allá de la cadena de suministros, lo que está en juego es la capacidad de proveer cables, tubos, satélites, fibra óptica, y todo tipo de infraestructura que garantice que los datos estén en manos de las potencias en disputa, y al mismo tiempo, posibiliten que el desarrollo de la IAGen no se tope con obstáculos materiales que amenacen el crecimiento exponencial en el poder de cómputo.

En sexto lugar, se destacan una variedad de *minerales* y materiales raros requeridos para la producción de los componentes electrónicos del hardware e infraestructura⁷. En esta capa, no solo intervienen empresas multinacionales y Estados, sino también cooperativas y formas organizativas no institucionalizadas que llevan adelante minería ilegal en diferentes países

⁶ Nos referimos no solo a los dispositivos y soportes con los que interactuamos con la IAGen, sino también a los semiconductores y los componentes electrónicos que garantizan el poder de cómputo de las tecnologías digitales

⁷ Incluye desde el cobre, litio, zinc, cobalto, hasta otros minerales como tungsteno, tantalio, indio, galio y las llamadas tierras raras.

como la República Popular del Congo, Brasil, y Perú. Así como gran parte de la mano de obra involucrada en el desarrollo de los algoritmos permanece invisible, las actividades extractivas que operan en estos contextos, también son ocultadas.

La séptima capa es la que permite que todas las demás puedan operar. Comprende a toda la actividad de extracción de *energía*⁸. La infraestructura que sostiene a la IAGen es altamente demandante de energía. Los bits mueven el mundo, pero los átomos lo componen. En esta capa, están involucradas las grandes empresas energéticas privadas y públicas, pero también se identifica una fuerte intervención de los Estados que promueven, por ejemplo, políticas de descarbonización e incentivo a la transición energética.

Descifrando el iceberg de la IAGen

Los bits no son solo bits: una forma de codificación del poder

Destacar que los bits *no solo* son bits implica problematizar que toda tecnología digital conlleva en su código técnico los criterios y reglas implícitos y explícitos que codifican y/o modulan las posibilidades y restricciones de instalación, uso y apropiación (Lessig, 2006). Es decir, lejos de ser “neutrales”, las tecnologías son inherentemente políticas (Winner, 1985), recuperando sesgos y parámetros presentes en el diseño pero también generando posibilidades inesperadas en el mismo; co-proyectando y co-diseñando el mundo.

En segundo lugar, los sistemas de perfilado algorítmico - presentes en la mayoría de las acciones utilizando objetos técnicos digitales (ODT), principalmente en las plataformas digitales, cada vez más se van posicionando como co-artífices de la realidad por su ejercicio de recomendación y modulación continua de deseos, subjetividades y pautas para la acción, siendo asimismo mecanismos de extracción masiva de datos y de atracción de la

⁸ Comprende tanto las fuentes renovables (energía eólica, solar, hidráulica, nuclear, etc.) como no renovables (gas y petróleo convencional y no convencional).

atención (Sadin, 2020). Sumado a esto, el ser creados mayormente bajo software privativo y secretos industriales, se les otorga un carácter mayormente opaco, *cajanegrizando su composición y operatoria* (Finn, 2017), dificultando su auditoría por parte de individuos, organizaciones de la sociedad civil y hasta de los mismos Estados (Gendler, 2023b). Lejos de ser “tecnologías neutrales”, los sistemas algorítmicos se asientan como mecanismos opacos de ejercicio del saber-poder, codificando y sugiriendo en base a la conjunción de sus normas de diseño, los datos extraídos y los objetivos político-económicos de las empresas que los crean (Rodríguez, 2019).

Los principales desarrollos de la IA y los que competen a la IAGen profundizan estas lógicas y características. Al estar imbricados con la actual fase del proceso de datificación social (Gendler, 2024a), ser producidos por el mismo conglomerado de empresas que despliegan las plataformas digitales, y al sostener el carácter opaco mencionado anteriormente, es posible postular que las problemáticas previamente mencionadas se continúan en los LLM y otros desarrollos de IAGen, e incluso, por sus características disruptivas, encuentran en ellos un vehículo para lograr un nuevo salto de escala (Gendler & Mallamaci, 2024). Asimismo, las prácticas más recientes de incorporación de los LLM en las principales plataformas digitales evidencian una estrategia de integración a las mismas⁹. Esto incrementa su utilización y mejora como también plantea nuevos escenarios en la economía política de la IAGen y el ecosistema de plataformas (Lassi, 2025).

Desde las empresas creadoras y otros aliados tecnoentusiastas se le suele brindar a los procesos de los desarrollos de IA un *carácter de superioridad* en base a su tratamiento automatizado de datos que “reflejan la realidad” y “libre de sesgos”, postulándolas por encima de las decisiones orgánicamente humanas y alentando a seguirlas, a la vez de abrir la puerta al cuestionamiento de toda acción y/o decisión que no lo haga (Innerarity, 2024).

⁹ Un claro ejemplo de esto es la presencia cada vez más central del LLM Gemini a la hora de utilizar las búsquedas de Google.

Trabajo cognitivo invisible: ¿qué tan artificial es la IA?

Al mirar el sistema global de la IAGen se identifican un conjunto de tareas que suelen ser poco conocidas y que cuesta identificar con precisión. Se trata de un trabajo cognitivo que resulta vital para el entrenamiento de los algoritmos de la IAGen y que en algunos casos puede implicar tareas rutinarias que requieren baja calificación (Reynolds et al., 2024).

Gray (2019) y Pasquinelli (2022) caracterizan a este como *trabajo fantasma* o *invisible*, dado que resulta poco conocido y se lo suele ocultar, dado que no representa la labor cognitiva asociada al diseño de algoritmos, sino que se trata de microtareas que no siempre requieren de una alta calificación. Asimismo, para su realización se conforman redes de *crowdsourcing* (Schmidt et al., 2020), integradas por trabajadores humanos mal remunerados (Berg, 2018) cuyas tareas muchas veces les generan daños psicológicos (Roberts, 2019). Este trabajo remite a tres *microtareas*, las cuales comprenden acciones específicas, acotadas, y repetitivas que se reparten entre muchos trabajadores, asignadas de forma automatizada, cuyo objetivo remite a perfeccionar el proceso de desarrollo de los algoritmos.

En primer lugar, se destaca el *labelling*. Refiere al etiquetado de datos para entrenamiento de los algoritmos. La tarea supone identificar datos sin procesar (imágenes, videos, texto) y luego añadir una etiqueta para especificar su contexto, permitiendo mejorar la predictibilidad del modelo de *machine learning* que se está desarrollando.

En segundo lugar, sobresale el *scoring*. Este constituye el perfeccionamiento de los algoritmos a partir de una serie de procedimientos. El proceso se realiza sobre los resultados del modelo de *machine learning*, y no sobre su insumo únicamente, tratando de entrenar al algoritmo, para verificar si su comportamiento es el indicado, o modificar algunos de sus hiperparámetros. En comparación a la primera tarea, se trata de una labor que suele requerir mayor calificación (conocimientos respecto a la ciencia de datos, por ejemplo).

Por último, se destaca la *moderación de contenido*. Refiere a la detección de sesgos en las bases de datos o información problemática como la circulación de contenido violento o discriminatorio en redes sociales. Si bien los algoritmos comienzan a detectar algunos de estos sesgos, todavía hace falta asistencia humana que realizan determinados trabajadores. Aquí las competencias laborales se centran en habilidades asociadas a la detección del contenido que se aparta de los marcos legales que encuadran la actividad, como también del que puede resultar hiriente para otros usuarios al exceder lo aceptable por parte de las plataformas. En este tipo de tareas, dada la exposición a contenido violento y perturbador, se ponen en juego aspectos emocionales y psicológicos que afectan seriamente a los trabajadores (Chen, 2017). Visibilizar el trabajo detrás de los algoritmos, permite refutar el aparente *carácter artificial* de la IA.

Datificación, plataformización, y colonialismo digital

A la hora de problematizar respecto de los datos, los cuales constituyen la materia prima digital esencial de todo el sistema global de la IAGen, y su procesamiento, es necesario comprenderlos dentro de un proceso de mayor alcance y duración: la datificación social. Esta refiere a los procesos donde un elemento concreto o simbólico es cuantificado y registrado, convirtiéndose en datos y, por tanto, implica técnicas de recolección, medida, almacenamiento para usos teóricos o aplicados (Mejías & Couldry, 2022).

Lejos de ser un hito reciente, la datificación social es un proceso que presenta diversas etapas (Krotz, 2024), desde sistemas de medición del tiempo y el espacio, cantidades, clima, tamaños, principalmente para la administración urbana y la logística militar, hasta, más avanzado en el tiempo, su aplicación a gran escala para la gestión de las poblaciones humanas (Foucault, 2007).

Recuperando lo trabajado en producciones previas (Gendler, 2021; 2023b; 2024a), es posible identificar como principales saltos de escala de la datificación social a la creación de ARPANET y la revolución de los microprocesadores a partir de 1970; la masificación de Internet y el

nacimiento de la web a inicios de la década de 1990; la crisis de la “burbuja punto com” a inicios del nuevo milenio; el surgimiento del *smartphone* en conjunto con la crisis económica de 2008; la introducción quirúrgica de modelos de IA en las plataformas digitales para optimizar los mecanismos algorítmicos de recomendación y personalización (modelo RAPPa); la pandemia del COVID-19 y la digitalización de la experiencia durante esa particular etapa; y la masificación de la IAGen tras la puesta a disponibilidad de ChatGPT 3.5 en noviembre de 2022. Todos estos hechos, y otros menores que han acontecido durante ese lapso de tiempo, permiten identificar saltos cuantitativos y cualitativos de la disponibilidad y mecanismos de extracción y almacenamiento de datos para diversos procesos, entre ellos el entrenamiento de los modelos generativos. Asimismo permiten ahondar en la variabilidad de distintos tipos de datos en función de la fuente de la que provengan.

En primer lugar, se consideran los *datos provenientes de Internet*, los cuales principalmente son recolectados en los “terrenos” de las plataformas digitales. También es posible considerar aquí las bases de datos públicos -en crudo- provistas por los gobiernos, especialmente bajo los planes de transparencia de Datos Abiertos.

En segundo lugar, *datos biométricos y de salud*, tanto los que provienen de registros públicos -huella digital, iris, imágenes de rostro-, junto a otros adquiridos vía dispositivos *wearables* -pulsaciones, oxígeno en sangre, frecuencia de sueño, entre otros-.

En tercer lugar, *datos del entorno*, tanto rural, urbano y/o climáticos, adquiridos vía redes de sensores públicos y/o privados para registrar y generar en cuasi tiempo real informes de estado de situación y recomendaciones de acción, ya sea para la siembra, irrigación, gestión urbana, entre otras. Es posible considerar también aquí los datos obtenidos del entorno hogareño.

En cuarto lugar, *datos provenientes de fuentes de operación humana-maquinica*, especialmente los obtenidos en los procesos industriales gracias a la red de sensores y conectividad 5G. Es posible registrar datos de operaciones de

producción, cantidades y distribución de trabajadores, materias primas y productos, sugerencias logísticas, entre un enorme etcétera.

En quinto lugar, *datos complejos ya procesados*, mayormente disponibles en Internet y otras fuentes digitalizadas como artículos de prensa, libros, informes técnicos, *papers* científicos, entre otros. Si bien este tipo es similar al primero, la diferenciación remite a que los datos obtenidos de la web a secas son datos “en bruto” mientras que los aquí referidos ya contienen procesos de iteración, correlación y, muchas veces, análisis.

Tabla 3

Tipos de datos

Tipos de datos	Características	Mecanismos extractivos	Actores dominantes
Provenientes de Internet	Huella digital. Consumo de plataformas. Datos producidos por usuarios.	Datificación y plataformización. Generar, almacenar, procesar y gestionar grandes volúmenes de datos.	Gigantes tecnológicos
Biométricos y de salud	Datos biológicos que remiten a imágenes de rostro, información genética, huella dactilar, iris, entre otros.	Extracción por parte de gobiernos a través de diferentes medios: escaneo de iris, cámaras, sensores, extracción de ADN, entre otros. Extracción directa por parte de empresas que monetizan esos datos.	Estados; gigantes tecnológicos; empresas multinacionales
Del entorno	Datos y métricas de procesos urbanos (tráfico, niveles cloacales, alumbrado) y rurales.	Extracción por redes de sensores y tecnologías de IoT ubicadas en espacios rurales y ciudades.	Estados; empresas multinacionales
Provenientes de fuentes de operación humano-maquinica	Datos de procesos industriales, de producción de materiales y de localización de maquinaria.	Extracción por redes de sensores y dispositivos IoT industriales.	Estados; empresas multinacionales
Información y conocimiento	Datos ya procesados de alguna manera: artículos de prensa, libros, papers científicos, repositorios de código, entre otros.	Extracción por parte de empresas de la información y el conocimiento de acceso libre en Internet. Compra de bases de datos.	Gigantes tecnológicos; empresas multinacionales

Nota. Elaboración propia.

Asimismo, la excesiva cuantificación de la vida añade nuevos desafíos decisionales en torno a la capacidad de autonomía pero también de responsabilidad de las elecciones tomadas (Innerarity, 2024). Finalmente, la extracción de datos complejos ya procesados despliega una nueva problemática acerca de la propiedad intelectual y el usufructo privado de las mismas, especialmente aquellas generadas originalmente con fondos públicos.

La batalla por el hardware

En esta capa, se despliega una compleja y poderosa cadena de suministro que reúne a grandes empresas, las dos potencias en disputa y sus Estados subordinados, en un esquema que está sometido a presiones constantes.

En el campo de los dispositivos y componentes electrónicos, la disputa entre las dos superpotencias del momento, EEUU y China, se materializa en tres áreas específicas. En primer lugar, la producción de semiconductores representa un punto nodal de la soberanía tecnológica y otorga ventajas en materia de defensa. Las estrategias de *reshoring* desplegadas por EEUU como la sanción de la ley del Chip (*CHIPS and Science Act*) bajo el gobierno de Biden¹⁰, pueden ser entendidas como síntomas de un recrudecimiento de la disputa por la hegemonía global, buscando reducir la interdependencia entre las potencias en pugna.

En un segundo plano, se identifica la presencia de corporaciones tecnológicas de la potencia enemiga en el territorio como un asunto que pone en riesgo la propia institucionalidad, seguridad e intereses nacionales. La expulsión de la empresa china Huawei y la presión para la venta de TikTok en EEUU, son ejemplos de esta mirada.

Por último, a raíz de la escalada en el enfrentamiento entre las potencias y su impacto en el escenario internacional, se interpreta a la

¹⁰ Se trata de una ley de incentivo a la producción doméstica de semiconductores a través de fondos específicos de fomento a la actividad industrial y de investigación en este campo. Accesible en: <https://acortar.link/FQPVVM>

producción doméstica del hardware como un elemento indispensable de autonomía y fortaleza. Similar a lo que sucedió con la pandemia, los países que tuvieron la posibilidad de autoabastecer la producción doméstica y los materiales e insumos que garantizan el acceso a los bienes básicos, contaron con una ventaja. En el sistema global de la IAGen, el hardware puede constituir un limitante a las posibilidades de desarrollo del software, por ende, resulta esencial aumentar el control sobre su producción.

Sin embargo, en la producción de hardware, una fase en la cadena de producción y consumo se suele ocultar: los desechos electrónicos. Se trata de una de las grandes externalidades que este sistema descarga sobre los países periféricos. De acuerdo con el Observatorio Mundial de los Residuos Electrónicos (2020)¹¹, China y EEUU son los mayores generadores de chatarra electrónica per cápita y, junto a la UE, son los mayores consumidores de productos electrónicos a nivel mundial, pero tan solo reciclan el 15%, el resto va a parar a los vertederos de países en vías de desarrollo, la mayoría en África¹². La paradoja es que en esos países, recolectores remueven la basura para buscar minerales que esos desechos electrónicos contienen, muchos de ellos originalmente extraídos en esos mismos países.

Otro punto a atender son las condiciones laborales en las fábricas en las que se producen dispositivos tecnológicos, dado que la gestión algorítmica también afecta la producción y comercialización de bienes físicos. La elevada tasa de suicidios en trabajadores chinos en Foxconn, el mayor fabricante de Apple, sumado a una situación similar entre trabajadores norteamericanos de Amazon, constituye uno de los tantos ejemplos.

Desmalezar la nube

La infraestructura constituye una pieza esencial para articular la capa correspondiente al hardware (los dispositivos tecnológicos a través de los cuales se interactúa con la IAGen) con el software (los propios algoritmos y

¹¹(Forti et al., 2020)

¹² Accesible en: <https://acortar.link/bTvmtM>

la interfaz digital), componiéndose asimismo de un ensamble de hardware y software. Por un lado, comprende una variedad de componentes electrónicos con los que se construyen los centros de datos (GPU, Racks, servidores, sistemas de refrigeración, etc.), y una diversidad de dispositivos que permiten conectar los dispositivos a una red (sensores, módems, satélites, entre otros). Por otro lado, incluye el software que permite “virtualizar” parte de los recursos físicos mencionados anteriormente y conectar el funcionamiento de los componentes electrónicos al entorno de la nube, de manera que el proceso de almacenamiento y gestión de datos incremente su grado de automatización.

Se puede dividir la capa de la infraestructura material que sostiene la producción y consumo de los bienes intangibles en dos campos. Por un lado, el almacenamiento de los datos que constituyen la materia prima digital más relevante del sistema. Por el otro, las redes de conectividad digital que posibilitan que los datos sean conducidos a los centros de almacenamiento de datos, pero también que puedan circular por toda la red. En ambos casos, el desafío es incrementar la capacidad de almacenamiento y la velocidad de transmisión, circulación, y generación de datos.

En relación con el almacenamiento de datos, en Occidente tres de las cinco GAFAM, Amazon Web Services, Google Cloud y Microsoft Azure, disponen de los centros de datos más grandes a nivel mundial, hegemonizado la computación en la nube. Por el lado de China, se destacan Baidu Cloud, Huawei Cloud, Alibaba Cloud, y Tencent Cloud que se reparten el mercado de la nube en el gigante asiático.

Ahora bien, la “nube” constituye un ensamble de hardware y software que permite almacenar y acopiar un gran volumen de datos, para que luego se puedan procesar y gestionar. Los datos allí almacenados, presentan un atributo físico/material, dado que se requiere de grandes servidores que, a su vez, necesitan de un conjunto de aparatos y componentes electrónicos para funcionar, pero también disponen de un atributo esencialmente virtual/digital, porque son datos atravesados por el tamiz digital (son bits).

Además de un ensamble sociotécnico específico que requiere de grandes inversiones de capital y de una movilización de recursos tangibles e intangibles a escala global, la nube constituye un servicio que permite que la información que millones de usuarios, empresas, y gobiernos generan, día a día, no tenga que ser almacenada en sus propios dispositivos. Se trata de un modelo de negocios en el cual se vende el acceso a los recursos de la infraestructura (Gendler, 2021).

Uno de los puntos centrales se vincula con la propiedad y uso de los datos que allí se almacenan. Dado que los usuarios, ya sean individuos, empresas, y/o gobiernos, depositan sus datos en una infraestructura que no les pertenece, se instala una controversia respecto a si continúan siendo propietarios de los datos que generan, o por el contrario, al momento de exteriorizar su almacenamiento, también aceptan un procedimiento de desposesión de esos datos. Si bien, continúan teniendo acceso a los mismos, y rigen derechos de propiedad intelectual sobre ellos, las empresas dueñas de los centros de datos donde se alojan, pueden disponer de ellos para otros fines, como por ejemplo, entrenar modelos de IAGen. Las mismas empresas que proveen los servicios de la nube son las que desarrollan varios de los principales desarrollos de IAGen, entre ellos, los LLM, por ejemplo, Google (Gemini), o establecen alianzas con las empresas que desarrollan esos modelos, como Microsoft con OpenAI.

Respecto a la conectividad, la infraestructura más relevante en la etapa actual se asocia al 5G. Algunas de las empresas que dominan la provisión de infraestructura de red, como Ericsson, CISCO, Qualcomm, Nokia, y Huawei, distribuyen sus operaciones entre América del Norte, Europa, América Latina, y Asia y Pacífico. Sin embargo, a raíz de la disputa por la hegemonía global entre EEUU y China, es posible diferenciar dos ecosistemas de 5G correspondientes a las potencias en pugna, cada uno con su propia red de actores, pero también de normas, licencias, protocolos de red y capitales (Borrastero, 2021). En la actualidad, China registra el máximo número de pedidos de patentes vinculadas al 5G a nivel mundial con 28.303, seguido por

EEUU con 21.791, los dos muy por encima del tercer país, Corea del Sur, con 12.930¹³.

A diferencia de las generaciones previas, el 5G conlleva dos nuevas características: reducción de la latencia a valores insignificantes y la posibilidad de interconexión de miles de dispositivos en simultáneo. Si bien esto suele ser destacado en torno a la implementación de prácticas y consumos en tiempo real, como las ciber-cirugías, también abre la puerta a permitir una aceleración de los procesos de extracción de datos, a optimizar la recolección de los datos del entorno y/o la implementación de aplicaciones que generen un mayor caudal de bits, como también la priorización de los flujos de datos (Gendler, 2024b), utilizándose para ello desarrollos de IA (Bazarena García et al., 2022).

Detrás de la conectividad que garantiza el 5G, emergen una variedad de dispositivos electrónicos que constituyen su soporte material: antenas, radares, sensores, satélites, entre otros. Además, se requieren cables submarinos, millones de kilómetros de tendido de fibra óptica, tubos de aluminio, entre otros. Esto quiere decir que detrás de esa infraestructura, emerge una cadena de suministros que opera a nivel global que involucra a la industria pesada. Una paradoja del industrialismo que no termina de morir.

Asimismo, la masificación del 5G abre la puerta a nuevos desafíos en torno a la privacidad de los datos, en una mayor escala de aceleración debido a sus características distintivas. Debido a esto, sumado a la concentración de las infraestructuras en pocas corporaciones, y su uso potencial para fines que no fueron avalados por los propios usuarios, es que la conformación de una nube pública, que cada Estado pueda desplegar y sostener de manera autónoma y soberana en articulación con otros actores, se torna un desafío ineludible en el presente.

¹³ Accesible en: <https://www.gov.br/inpi/pt-br>

Los bits mueven el mundo, los átomos lo componen

La materialidad subyacente a la IAGen se expresa en la materia prima requerida para fabricar los componentes electrónicos y dispositivos tecnológicos de las capas del hardware y la infraestructura. Se trata de toda una variedad de minerales, materiales y tierras raras que se extraen en diferentes territorios para abastecer una cadena de suministros global que enlaza los bienes informacionales con los bienes materiales.

Como fue mencionado en el segundo apartado, la IAGen constituye una tecnología netamente extractiva. Su desarrollo requiere, de manera permanente, intensiva y a una gran velocidad, expandir la frontera de extracción sobre la propia capa geológica del planeta (Estampa, 2024). Según Gabrys (2013) para producir un microchip de memoria de dos gramos se necesitan 1.3 kilos de combustibles fósiles y materiales, los cuales el 99% se desechan en el proceso de producción. La cadena de suministros que permite proveer una gran parte de estos elementos se desenvuelve en un marco de opacidad y falta de transparencia respecto al origen de algunos de esos insumos. No se trata solamente de comprender cómo se transforma el silicio en un chip, sino dilucidar un entramado denso y capilar de redes materiales que abarcan la totalidad del planeta.

De esa variedad de minerales y materiales que forman parte de la cadena global de producción de hardware y construcción de la infraestructura digital, algunos adquieren especial relevancia en el sistema global de la IAGen. La gran mayoría se concentra en los países periféricos, aunque China y EEUU están avanzando fuertemente en la extracción de algunos de estos en su propio territorio¹⁴.

¹⁴ A modo de ejemplo, el gigante asiático constituye el primer productor mundial de tierras raras. Datos disponibles en RTVE (2025). En el caso norteamericano, la extracción de litio en el Estado de Nevada, avanza a paso firme. Si bien no alcanza para cubrir la demanda, y EEUU no forma parte de los grandes productores de este mineral, a diferencia de China, que ocupa el tercer lugar, en 2023 se encontró una de las reservas más grandes del mundo, lo que, eventualmente, podría modificar esta ecuación (Bazan, 2025).

En primer lugar, el cobre y el oro conforman una pieza significativa del engranaje material de la IAGen. Su conductividad eléctrica los transforma en minerales esenciales tanto para la infraestructura digital como para la producción de hardware. Chile y Perú constituyen los dos mayores productores de cobre a nivel mundial. China, su principal comprador. En el caso del oro, China y Australia constituyen los dos mayores productores a nivel mundial, seguido de Rusia, Canadá y Estados Unidos. Sin embargo, la minera ilegal ubicada en la Amazonia brasileña constituye uno de los mayores proveedores para la industria del hardware (Manzoli et al., 2021).

En un segundo plano, aparece el litio, mineral estratégico a partir de constituir el principal componente de las baterías eléctricas. En el marco de la transición energética, pero también como componente esencial de los dispositivos digitales, este mineral resulta primordial en la estructura global de la IAGen. Australia es el mayor productor mundial de litio, seguido de Chile y China. Argentina y Bolivia, si bien producen una cantidad mucho menor, disponen de las mayores reservas a nivel mundial, dado que junto a Chile, conforman el “triángulo del litio” (Bazan, 2025). Estas actividades productivas generan fuertes tensiones con las comunidades locales, por el consumo de recursos escasos y también por el impacto que generan sobre el ambiente (Grieco, 2023).

Un cuarto mineral esencial en la producción de tecnologías digitales, y en las baterías eléctricas, es el cobalto. La República Popular del Congo constituye el principal productor mundial de este mineral, muy por encima de sus competidores (Gerding, 2024). ONG internacionales que se avocan a la defensa de los derechos humanos como Amnesty Internacional, denunciaron que la extracción de este mineral en el Congo incluye una variedad de abusos y situaciones de violencia contra las comunidades locales que incluyen desalojos forzados, trabajo infantil e incluso agresión sexual (Aministia Internacional, 2023). Tanto en empresas multinacionales como por los grupos armados y algunas pequeñas cooperativas, la minería en este país de África se lleva adelante de manera artesanal: trabajo manual que utiliza herramientas básicas y una tecnología precaria.

Además de los señalados, se destaca el creciente peso que tiene la producción tanto de tierras raras como una variedad de materiales como tungsteno, tantalio, indio, y galio, entre otros. Estos últimos se encuentran repartidos por todo el planeta, pero su extracción es compleja e implica, en algunos casos, un gran impacto ambiental. Por esta razón, la mayoría de los yacimientos activos se localizan en los países periféricos.

Detrás de los bits que mueven el mundo subyacen los átomos que lo componen. Según diversos informes, como el del programa de Naciones Unidas para el Medio Ambiente, la creación, masificación y penetración de la IAGen crecientemente plantea desafíos ecológicos en diversos planos (ONU, 2025). El creciente costo en materias primas de la producción del hardware necesario para su funcionamiento y masificación constituye un tema ineludible en el debate en torno a la ética de la IA. No solo por el impacto sobre el medio ambiente y las problemáticas ecológicas asociadas al extractivismo, sino también por los conflictos y tensiones sociales y políticas que atraviesa la producción de estos materiales (Dauvergne, 2020).

Es la energía, estúpido

La energía es la capa que alimenta a todas las demás. Sin energía, ningún nodo de la red puede funcionar. Comprende las fuentes renovables (energía eólica, solar, hidráulica, nuclear, entre otras) como no renovables (fundamentalmente gas y petróleo convencional y no convencional). Una particularidad del sistema global de IAGen es que demanda y consume grandes cantidades de energía. A modo de ejemplo, el uso de los LLM como ChatGPT o Gemini, consumen diez veces más electricidad que una búsqueda de Google, con lo que la masificación y reemplazo que se viene proyectando también generaría un impacto ambiental significativo (Jimenez, 2024). En los estados norteamericanos en los que se localizan los grandes centros de datos de ese país, se estima que más de un 10% del total de la energía consumida responde a este sector (IEA, 2024).

La producción de energía se encuentra distribuida geográficamente a lo largo y ancho del mundo, pero se destaca la concentración de grandes

reservas de gas y petróleo convencional y no convencional en determinados países como los integrantes de la OPEP¹⁵ Rusia, Sudáfrica, Kazajistán, Argentina, México, Brasil, Estados Unidos, Canadá y China.

En relación con la transición energética, se destacan los países escandinavos que lideran esta industria junto a Francia, Alemania, y Austria¹⁶. En el sector energético, predominan las empresas multinacionales pero también participan activamente empresas de capital estatal, gobiernos y una red extensa y heterogénea de organizaciones de la sociedad civil interesadas en la transición energética. También hay que destacar la Conferencia de las Naciones Unidas sobre el Cambio Climático (COP29) que constituye la agenda global de carácter multilateral más relevante en materia de compromisos con la lucha contra el cambio climático, en la cual el debate por la energía ocupa un rol trascendente (ONU, 2024).

La energía como capa nodal del sistema global de la IAGen está atravesada por tres desafíos que probablemente definan el carácter sostenible, en términos ambientales y ecológicos, de toda esta estructura. En primer lugar, dado que todo el sistema demanda más energía, emerge el interrogante de cuál será la fuente principal de esa energía. En la actualidad, son los combustibles fósiles los que abastecen al sistema¹⁷. ¿Puede la energía renovable modificar esta ecuación?

En un segundo plano, se introduce un debate en torno a la posibilidad de desarrollar con mayor intensidad los sistemas descentralizados que acercan la generación con el consumo de energía, dentro de los cuales, las fuentes renovables como la energía solar se posicionan como una alternativa posible.

¹⁵ La Organización de Países Exportadores de Petróleo (OPEP) está conformada por 12 países de África, Asia y América.

¹⁶ Se toma como fuente de información, el informe del Foro Económico Mundial (2024) y el informe elaborado por la Administración de Información Energética de los Estados Unidos (EIA, 2015).

¹⁷ En la actualidad, todo el sector TIC representa entre un 2 y un 4% del total de emisiones de carbono. Sin embargo, la huella de carbono de la IA puede incrementar significativamente estos valores sino se modifica la matriz energética (Pasek, 2023).

Por último, los costos de la energía. El desarrollo de la IAGen implica una inversión millonaria de capital para incrementar la base de datos con los que se entrenan los modelos, así como también producir el hardware y la infraestructura capaz de sostener este proceso en términos de capacidad de almacenamiento y procesamiento de la información¹⁸. En este contexto, el precio de la energía se torna un factor que o bien puede favorecer el desarrollo de todo el sistema global al proporcionar energía barata y por ende contribuir a reducir un componente de los costos, o puede producir el efecto contrario. En términos ambientales, la energía barata supone incentivos para incrementar su consumo, más allá de las innovaciones en el campo de los sistemas descentralizados y la energía limpia, mientras que la energía cara, podría resultar un factor catalizador para acelerar la transición energética.

Conclusiones

En el presente artículo se ha realizado un amplio recorrido por las características, problemáticas, actores y desafíos de la IAGen a nivel global, entendiéndola como una de las principales tecnologías disruptivas dentro de la segunda fase del informacionalismo, el paradigma 4.0.

En este sentido, se han definido y abordado las características y problemáticas de la IAGen, para luego dar paso a la presentación del modelo de siete capas, considerando los niveles de incumbencia, las tecnologías y recursos técnicos y naturales involucrados, los principales países y actores implicados, las empresas y otros actores relevantes.

Asimismo, cada capa ha sido descrita y problematizada con gran nivel de detalle en torno a sus características distintivas, las problemáticas generales que presentan, las relaciones entre los distintos niveles macro, meso y micro que implican. Se espera que esta elaboración pueda ser una herramienta analítica y a la vez conceptual que exceda el ámbito académico. En este sentido, se considera que esta propuesta también puede permitir ahondar en un conocimiento integral y a la vez crítico que permita a tomadores de

¹⁸ El CEO de DeepMind, compañía que adquirió Google en 2024, afirmó que probablemente inviertan más de 100 billones de dólares en IA (Connaster, 2024).

decisiones, legisladores, organizaciones de la sociedad civil, y consultores, entre otros, optimizar los recaudos y estrategias respecto de oportunidades y riesgos a la hora de evaluar posibles regulaciones, normativas, prácticas, políticas públicas y procesos de transformación y/o implementación institucional.

Queda planteada aquí la sistematización y las posibles recomendaciones por seguir.

Será cuestión de futuros trabajos el profundizar sobre cada una de las capas aquí descritas para poder ahondar en los diversos conflictos y niveles que conlleva al despliegue de una tecnología digital disruptiva que solamente ha empezado a mostrar sus primeros pasos.

Para futuras investigaciones quedan planteados una serie de interrogantes que se tornan determinantes para pensar cómo se reproduce un sistema global que tiene cada vez mayor incidencia en la forma en que se estructura el poder, el trabajo y la generación de riqueza en el mundo: ¿Cómo se expresa este sistema global sobre los territorios y comunidades? ¿Qué nuevas jerarquías se producen? ¿Vamos camino a una hipercentralización del poder y la riqueza? ¿Cómo se distribuyen las externalidades ambientales, sociales, y políticas que produce el desarrollo de la IAGen? ¿Se profundiza el colonialismo digital?

Referencias

- Amnistía Internacional. (2023, 12 de septiembre). *República Democrática del Congo: La minería industrial de cobalto y cobre para baterías recargables está dando lugar a graves abusos contra los derechos humanos*. <https://www.amnesty.org/es/latest/news/2023/09/drc-cobalt-and-copper-mining-for-batteries-leading-to-human-rights-abuses/>
- Bazan, M. (2025, 15 de febrero). Conozca el ranking de los principales países productores de litio en 2024. *Rumbo Minero Internacional*. <https://www.rumbominero.com/peru/noticias/internacionales/conozca-a-los-principales-paises-productores-de-litio-en-el-mundo/>

- Belzarena García, P. J., Larroca Ponzoni, F., Bermolen Romeo, M. P., Inglés Loggia, L., Belcredi Zambra, G., & Randall Carlevaro, M. (2022). *Informe final del proyecto: Inteligencia Artificial aplicada a redes 5G*. Agencia Nacional de Investigación e Innovación (ANII). <https://hdl.handle.net/20.500.12381/3200>
- Berg, J. (2018). *Digital labour platforms and the future of work: Towards decent work in the online world*. International Labour Organization.
- Borrastero, C. (2021). *5G y territorialización de la disputa tecnológica global: EEUU, China, América Latina*. [Ponencia]. 1º Jornadas CEUR - Tecnología y territorio. Paradigmas tecno-económicos y nuevas configuraciones socio-espaciales en América Latina, Buenos Aires, Argentina.
- Castells, M. (1999). *La era de la información: Economía, sociedad y cultura* (Vol. 1). Siglo XXI Editores.
- Chen, A. (2017, 28 de enero). The human toll of protecting the internet from the worst of humanity. *The New Yorker*.
- Connaster, M. (2024, 17 de abril). Google will pump more than \$100B into AI, says DeepMind boss. *The Register*. https://www.theregister.com/2024/04/17/google_deepmind_funding/
- Costa, F., Mónaco, J. A., Covello, A., Novidelsky, I., Zabala, X., & Rodríguez, P. (2023). Desafíos de la inteligencia artificial generativa. Tres escalas y dos enfoques transversales. *Question/Cuestión*, 3(76), e844. <https://doi.org/10.24215/16696581e844>
- Crawford, K. (2023). *Atlas de inteligencia artificial: Poder, política y costos planetarios de la inteligencia artificial*. Fondo de Cultura Económica.
- Crawford, K., & Joler, V. (2018). *Anatomy of an AI system: The Amazon Echo as an anatomical map of human labor, data and planetary resources*. AI Now Institute and Share Lab. <https://anatomyof.ai/>
- Dauvergne, P. (2022). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Review of International Political Economy*, 29(3), 696-718. <https://doi.org/10.1080/09692290.2020.1814381>
- EIA. (2015, 24 de septiembre). *World shale resource assessments*. U.S. Energy Information Administration. <https://www.eia.gov/analysis/studies/worldshalegas/>
- Estampa. (2024). *Cartografía de la inteligencia artificial generativa* (E. Tisselli & P. Duarte, Trads.). <https://acortar.link/T7GZsl>
- Feldman, P. (2024). La economía global de la inteligencia artificial generativa (IAG): Un abordaje desde la periferia. En *Libro de actas de las IV Jornadas Democracias y Desigualdades* (pp. 1-15). Universidad Nacional de José C. Paz (UNPAZ).
- Feldman, P., & Girolimo, U. (2021). La industria 4.0 en perspectiva latinoamericana: limitaciones, oportunidades y desafíos para su desarrollo. *Revista Perspectivas*

- de Políticas Públicas, 10(20), 459-491. <https://doi.org/10.18294/rppp.2021.3645>
- Finn, E. (2017). *What algorithms want: Imagination in the age of computing*. The MIT Press.
- Foro Económico Mundial. (2024, 21 de julio). Estos 4 países lideran la transición energética. *Transición Energética*. <https://es.weforum.org/stories/2024/06/estos-4-paises-lideran-la-transicion-energetica/>
- Forti, V., Baldé, C. P., Kuehr, R., & Bel, G. (2020). *The global e-waste monitor 2020: Quantities, flows and the circular economy potential*. United Nations University (UNU)/United Nations Institute for Training and Research (UNITAR) – co-hosted SCYCLE Programme, International Telecommunication Union (ITU) & International Solid Waste Association (ISWA).
- Foucault, M. (2007). *Nacimiento de la biopolítica: Curso en el Collège de France (1978-1979)*. Fondo de Cultura Económica.
- Gabrys, J. (2013). *Digital rubbish: A natural history of electronics*. The University of Michigan Press.
- Gendler, M. (2021). Mapeando la dataficación digital y algorítmica. Genealogía, estado de situación y nuevos desafíos. *In Mediaciones de la Comunicación*, 16(2), 17-33. <https://doi.org/10.18861/ic.2021.16.2.3166>
- Gendler, M. (2023a). Del objeto técnico al objeto técnico digital: apuntes y aportes desde (y a las) elaboraciones de Gilbert Simondon en un marco algorítmico. *Anacronismo e Irrupción*, 13(25), 109-133. <https://doi.org/10.14409/an.2023.25.e0005>
- Gendler, M. (2023b). De la cibernética al metaverso: una genealogía de características, transparencias y opacidades algorítmicas. *Disparidades. Revista de Antropología*, 78(1), e001b. <https://doi.org/10.3989/dra.2023.001b>
- Gendler, M. (2024a). Datificación social e inteligencia artificial: ¿hacia un nuevo “salto de escala”? *Resonancias. Revista de Filosofía*, 17, 121–141. <https://doi.org/10.5354/0719-790X.2024.74503>
- Gendler, M. (2024b). De la neutralidad de la red al 5G: Cambios en las dinámicas de poder del ecosistema de internet. *In Mediaciones de la Comunicación*, 19(1), 81–105. <https://doi.org/10.18861/ic.2024.19.1.3528>
- Gendler, M. A., & Mallamaci, M. (2024). Los bits (no solo) son bits: genealogía, dimensiones y actualidad de los estudios sobre plataformas digitales. *Revista Hipertextos*, 12(21), e079. <https://doi.org/10.24215/23143924e079>
- Gerding, J. (2024, 19 de marzo). Lo que esconde la minería del cobalto en el Congo. *DW*. <https://www.dw.com/es/lo-que-esconde-la-minería-del-cobalto-en-el-congo/a-68618192>
- González Monserrate, S. (2022). The cloud is material: On the environmental impacts of computation and data storage. *MIT Case Studies in Social and Ethical Responsibilities of Computing*. <https://doi.org/10.21428/2c646de5.031d4553>

- Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.
- Grieco Gaspar, G. (2023, 6 de marzo). Litio: La gran controversia del oro blanco. *Noticias UNSAM*. <https://noticias.unsam.edu.ar/2023/02/28/litio-la-gran-controversia-del-oro-blanco/>
- Hawkins, Z., Lehdonvirta, V., & Wu, B. (2025). *AI compute sovereignty: Infrastructure control across territories, cloud providers, and accelerators* [Manuscrito no publicado]. Social Science Research Network. <https://doi.org/10.2139/ssrn.5312977>
- Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799-819. <https://doi.org/10.1109/TAI.2022.3194503>
- IEA. (2024, 14 de enero). *Artificial intelligence*. International Energy Agency. <https://www.iea.org/topics/artificial-intelligence>
- Innerarity, D. (2024). *Artificial intelligence and democracy*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000389736>
- Jimenez, I. (2024, 12 de julio). La IA utiliza 10 veces más energía que el buscador de Google. *Las Empresas Verdes*. <https://lasempresasverdes.com/la-ia-utiliza-10-veces-mas-energia-que-el-buscador-de-google/>
- Krotz, F. (2024). Digitalisation today as the capitalist appropriation of people's mental labour. *tripleC: Communication, Capitalism & Critique*, 22(1), 208-231. <https://doi.org/10.31269/triplec.v22i1.1477>
- Lassi, A. (2025). Inteligencia artificial generativa integrada al ecosistema digital: Un marco de situación para la gubernamentalidad algorítmica. *In Mediaciones de la Comunicación*, 20(1). <https://doi.org/10.18861/ic.2025.20.1.3931>
- Lessig, L. (2006). *Code: Version 2.0*. Basic Books.
- Madakam, S., Holmukhe, R. M., & Revulagadda, R. K. (2022). The next generation intelligent automation: Hyperautomation. *Journal of Information Systems and Technology Management*, 19, e202219009. <https://doi.org/10.4301/S1807-1775202219009>
- Manzolini, B., Rajão, R., Haliuc Bragança, A. C., de Tarso Moreira Oliveira, P., Kenner de Alcântara, G., Nunes, F., & Soares Filho, B. (2021). *Legalidade da produção de ouro no Brasil*. Instituto de Geociências, Universidade Federal de Minas Gerais (IGC-UFMG).
- Mejias, U., & Couldry, N. (2022). Datificación. *Revista Latinoamericana de Economía y Sociedad Digital*, 1(1), 8-37. <https://doi.org/10.53857/hced6196>
- Naciones Unidas. (2024, 8 de octubre). *COP 29 is a vital moment for the world* [Video]. YouTube. <https://www.youtube.com/watch?v=CNI2SwiMcn0>
- Naciones Unidas. (2025, 13 de noviembre). La IA plantea problemas ambientales. Esto es lo que el mundo puede hacer al respecto. *Noticias ONU*. <https://www.unep.org/es/noticias-y-reportajes/reportajes/la-ia-plantea-problemas-ambientales-esto-es-lo-que-el-mundo-puede>

- Pasek, A. (2023). *How to get into fights with data centers: Or a modest proposal for reframing the climate politics of ICT* [Documento de trabajo]. Experimental Methods and Media Lab, Trent University.
- Pasquinelli, M. (2022). Cómo una máquina aprende y falla: Una gramática del error para la inteligencia artificial. *Revista Hipertextos*, 10(17), 13-29. <https://doi.org/10.24215/23143924e054>
- Reynolds, J., Quintero, D. F. P., Aguilar, J., & Kincaid, R. (2024). Work schedules, finances, and freedom: Work schedule fit and platform dependence among gig workers on Amazon's Mechanical Turk platform. *Socius*, 10. <https://doi.org/10.1177/23780231241286933>
- Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.
- Rodríguez, P. (2019). *Las palabras en las cosas: Saber, poder y subjetivación entre algoritmos y biomoléculas*. Cactus.
- Rosenberg, N., & Trajtenberg, M. (2004). A general-purpose technology at work: The Corliss steam engine in the late-nineteenth-century United States. *The Journal of Economic History*, 64(1), 61-99. <https://doi.org/10.1017/S0022050704002608>
- RTVE. (2025, 17 de enero). China encuentra su mayor depósito de tierras raras en el sur del país. *Agencia RTVE Noticias*. <https://www.rtve.es/noticias/20250117/china-encuentra-su-mayor-deposito-tierras-raras-sur-del-pais/16410876.shtml>
- Sadin, É. (2020). *La inteligencia artificial o el desafío del siglo: Anatomía de un antihumanismo radical*. Caja Negra.
- Schmidt, F. A. (2021). Mercados de trabajo digitales en la economía de plataformas: Mapeando los desafíos políticos del trabajo colaborativo y del trabajo de plataformas [gig work]. *Hipertextos*, 8(14), 11-58. <https://doi.org/10.24215/23143924e018>
- Schwab, K. (2016). *La cuarta revolución industrial*. Debate.
- Srnicsek, N. (2018). *Capitalismo de plataformas*. Caja Negra.
- UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial. https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa
- Valles, M. (1999). *Técnicas cualitativas de investigación social*. Editorial Síntesis.
- Vercelli, A. (2024). Regulaciones e inteligencias artificiales en Argentina. *In Mediaciones de la Comunicación*, 19(1), 105-135. <https://doi.org/10.18861/ic.2024.19.1.3549>
- Winner, L. (1985). ¿Tienen política los artefactos? En D. MacKenzie & J. Wajcman (Eds.), *The social shaping of technology: How the refrigerator got its hum* (pp. 55-67). Open University Press.
- Yuan, J. (2022, 9 de diciembre). The future is now: The AI gaming revolution. *Building Virtual Worlds*. <https://shorturl.at/PniQC>

Zukerfeld, M. (2020). Bits, plataformas y autómatas: Las tendencias del trabajo en el capitalismo informacional. *Revista Latinoamericana de Antropología del Trabajo*, 4(11), 1-50. <https://doi.org/10.35305/rat.v4i11.123>

**Religión y Nuevas Tecnologías:
Tecnología con alma: aportes y límites desde lo
religioso**
Religion and New Technologies:
Technology with soul: contributions and limitations
from a religious perspective

Alejandro Cerda Sanhueza

acerda@ucn.cl

Universidad Católica del Norte

ORCID ID 0000-0003-2009-1846

Resumen: El presente trabajo, pretende hacer una reflexión y propuestas a partir de la relación de las nuevas tecnologías, en especial las de comunicación, con la experiencia religiosa, tanto para vislumbrar algunos impactos y mutuas contribuciones, especialmente el aporte de la religión a la denominada cibercultura, que se constituye un espacio importante, donde se vive hoy en día lo religioso, y la existencia toda.

Los cambios culturales que se viven en el presente propiciados fundamentalmente dado por las nuevas tecnologías tienen impacto en toda la vida social y privada de las personas, como también de los diversos actores e instituciones que componen la sociedad. Esta nueva cultura tecnológica, se constituye en una oportunidad, un testimonio y un desafío para todos, en este caso particular para la religión, entendida esta como la entidad que relaciona lo humano con lo sagrado.

El trabajo, pretende enriquecer las miradas interdisciplinarias sobre el aporte y el impacto de la Inteligencia Artificial, y las nuevas tecnologías a los distintos ámbitos de la vida que se compone de diferentes dimensiones y perspectivas.

Palabras clave: Nuevas tecnologías; religión; antropología religiosa; cibercultura.

Abstract: This paper offers a reflective analysis and a set of proposals concerning the intersection between emerging technologies—particularly communication technologies—and religious experience. It explores the mutual influences and contributions between these domains, with special emphasis on how religion shapes and is shaped by the evolving landscape of cyberculture. Cyberculture has become a vital space where contemporary religious life unfolds, alongside broader aspects of human existence.

The cultural transformations driven by technological innovation are profoundly impacting both public and private spheres, as well as the institutions and actors that constitute society. This new technological culture presents opportunities, challenges, and testimonies for all sectors—especially for religion, understood here as the entity that bridges the human and the sacred.

The study aims to enrich interdisciplinary perspectives on the role and impact of Artificial Intelligence and other emerging technologies across diverse dimensions of life, offering insights into their implications for spiritual, social, and cultural practices.

Keywords: New technologies; religion; religious anthropology; cyberculture.

Introducción

*Nosotros los de entonces
ya no somos los mismos*

Pablo Neruda

Es evidente que los tiempos han cambiado, los nacidos en la década de los sesenta o setenta han sido testigo de saltos culturales, políticos, ambientales

y, sobre todo tecnológicos sin precedentes: han pasado del teléfono de disco a la nanotecnología, de la máquina de escribir antiguas a los procesadores de última generación, han visto como lo virtual, desplazan lo territorial, el uso del tiempo y espacio se han visto transformados, la virtualidad posibilita una forma de presencia que es, a la vez, ausencia.

La presencialidad se manifiesta de diversas formas, los ciclos virales del conocimiento, asociados al desarrollo psicobiológico y social, así como los respectivos protagonistas del proceso enseñanza - aprendizaje también se ven modificados, las relaciones son más simétricas entre formandos y formadores, las preguntas y respuestas fundamentales, han cambiado, las objeciones se han diversificado de tal manera que la incertidumbre se contrapone a la gran cantidad de datos que deambulan por las redes.

Los jóvenes, no ven a los mayores como mentores del conocimiento, en cuanto ellos tienen más información que los adultos y aprenden en todos los espacios, no sólo en las aulas. Por su parte los educadores buscan también reinstalarse en los nuevos escenarios del conocimiento y el denominado dataísmo (Harari, 2016).

La humanidad toda ha evolucionado tan significativamente en un breve periodo, que no se logra aquilatar la vertiginosidad, profundidad y constancia de dichos cambios. Es así como los cazadores recolectores, prontamente se transformaron en agrícolas, los pueblos primitivos poco demoraron en establecer la propiedad privada y el surgimiento del Estado y estructuras legislativas para proteger los bienes y la producción, siendo quizás esta la primera gran revolución cultural (Graeber & Wengrow, 2022). La técnica ha estado presente y ha sido el agente fundamental en la transformación sociocultural, se ha pasado del descubrimiento del fuego y la invención de la rueda, a la cuarta y quinta revolución industrial.

El cambio cultural más significativo, es quizás el nuevo habitat que se ha construido a partir del desarrollo científico, tecnológico, hoy habitamos la denominada cibercultura, ese “*universo de redes numéricas*” donde conviven sistemas técnicos, agentes culturales, prácticas simbólicas e instituciones en una trama híbrida y dinámica, dominada por las nuevas tecnologías y la

Inteligencia Artificial. Los algoritmos se apropian de la totalidad de la vida humana, y predice los gustos e instalan tendencias políticas, sociales y religiosas. Donde predomina el uso de forma ubicua de las nuevas tecnologías, Levy la describe en su obra *Cibercultura*:

Hay que tener en cuenta que la cultura digital es bastante más compleja e híbrida que los entramados de sistemas tecnológicos electrónicos y digitales que la configuran. Además de dichos sistemas materiales y simbólicos, están integrados en la misma agentes y prácticas culturales, interacciones y comunicaciones, colectivos, instituciones y sistemas organizativos, una multiplicidad de contenidos y representaciones simbólicas junto con los correspondientes significados, interpretaciones, legitimaciones, valores, etc. (Levy, 2007, p. 9)

Lo anterior tiene consecuencias antropológicas que se expresan en categorías tales como: pos-humanismo o transhumanismos. La primera hace alusión a una disolución en la que lo humano pierde protagonismo y se convierte en un aspecto más de un universo que no tiene principio ni fin, la segunda categoría es la proyección de lo humano por medio de las nuevas tecnologías, que lo harían casi inmortal y con poderes capaces de hacer todo lo que naturalmente no podría hacer (Perez Francisco, 2023). Estas nuevas tecnologías van configurando una forma de relación, de ser y estar, distinta a periodos anteriores de la historia, donde categorías como tiempo y espacio, sagrado-profano, interior - exterior, real y virtual, entre otras, se deconstruyen (Antolinos Asensio, 2019).

Este es el escenario en que se tiene que pensar al ser humano de hoy, de su forma de concebirse y de relacionarse. Las nuevas tecnologías están lejos de ser simples herramientas de uso, predominando así una concepción instrumental de la tecnología, son el espacio vital en el que se nace, se desarrolla y se existe, por lo tanto, son un contexto vital (Floridi, 2019).

Hasta el momento el ser humano, aún es protagonista de su historia y del desarrollo evolutivo de la especie, pero sin tener certeza de sí podrá seguir siendo su conductor, en cuanto el presente y el futuro está en gran

medida en manos de las nuevas tecnologías, que cada vez se van haciendo más independientes de la intervención humana, en sus procesos de aprendizajes y de creación de nuevas aplicaciones. Eso es algo inédito del cual se tiene que tomar conciencia, y por lo tanto se piensa poco desde las distintas disciplinas del saber (Yuval Noah, 2024).

Los cambios que van ocurriendo en este período tienen incidencias en las distintas dimensiones del ser humano, entendemos a este, como un ser, cultural, corporal, esencialmente comunicacional, y multidimensional; materia-espíritu, individuo, comunidad etc. (Perez Francisco, 2023). Estas dimensiones se complementan y requieren ser atendidas y desarrolladas, y para aquello el espacio que se habita es fundamental, y este nuevo escenario es el que ha ido modificando a la humanidad toda.

Desde una perspectiva humanista, una de sus dimensiones es su capacidad y búsqueda de trascendencia, de reconocerse como un ser finito y carente (de sentido, de infinitud, de vida eterna y super poderes), como lo afirmó Jaspers:

El hombre no es simplemente lo que es, sino lo que puede llegar a ser. Su ser está marcado por la carencia, por la conciencia de la muerte, el sufrimiento, la lucha. En estas situaciones límite, se revela la necesidad de trascender, de buscar sentido más allá de lo dado. (Jaspers, 1932)

Esa búsqueda de trascendencia lo impulsa a seguir buscando, explorando nuevos horizontes y transgrediendo fronteras de todo tipo (Boff, 2002). Se podría afirmar que es este mismo impulso el que lo ha llevado a encontrar en la tecnología un aliado para su búsqueda de infinitud y de satisfacción de necesidades. Y paradójicamente esta misma aliada que es la tecnología, ha ido afectando o resignificando esta dimensión. Por lo que la pregunta sería ¿de qué forma esta dimensión de trascendencia se ve afectada por las nuevas tecnologías?, en particular, la Inteligencia Artificial Generativa (IAG), ¿qué puede aportar esta dimensión a esta cibercultura?

El presente ensayo pretende hacer un esbozo de tensiones y reciprocidades entre lo religioso y las nuevas tecnologías. Plantea también la interrogante abierta qué aportes puede ofrecer la dimensión religiosa a este nuevo escenario (Cerda Sanhueza, 2020).

El itinerario de aproximación es, indagar sobre el surgimiento y configuración de la cibercultura y su concepción humanista. En un segundo momento, se pretende presentar la religión como un pulso antropológico de la era digital. A modo de conclusión, se expondrán algunas tensiones y mutuas contribuciones.

En resumen, no sólo los de entonces no son los mismos, como afirma el poeta chileno Pablo Neruda, sino lo que es más gravitante, es que el lugar que habita ese hombre o mujer es radicalmente diferente y con ello cambia su forma de entenderse, relacionarse y proyectarse.

La cibercultura y sus desafíos antropológicos

El verdadero peligro no es que las computadoras comenzarán a pensar como los hombres, sino que los hombres comenzarán a pensar como las computadoras.

Sydney J. Harris

La cibercultura ha trascendido los centros de investigación y los departamentos de informática para convertirse en un tema central de conferencias, foros empresariales y debates políticos. Universidades, start-ups, Ministerios de Salud y organismos de seguridad dedican recursos crecientes a comprender sus impactos. Este fenómeno es una interrogante no sólo a los especialistas, sino también para filósofos, juristas, teólogos y analistas sociales que buscan descifrar cómo las nuevas tecnologías, en especial las redes digitales transforman nuestras instituciones y prácticas cotidianas.

Este fenómeno exige de un estudio y diálogo interdisciplinar y del reconocimiento que se está ante un relato en construcción. No está todo

dicho, falta aún mucho por comprender (McKinsey Global Institute, 2023). Esto implica definir nuevas categorías y marcos conceptuales, para lo que se requiere de nuevas estructuras narrativas renovadas y flexibles.

Ella es un sistema integrado y complejo que involucra la manera como se producen, circulan y se reproducen un conjunto de prácticas culturales, acciones e interacciones. En este ecosistema digital, se homogenizan gustos y opiniones, cerrando espacios para la reflexión profunda y la alteridad (Levy, 2007). La velocidad de la información, la lógica algorítmica y la cultura de la inmediatez configuran una subjetividad marcada por eficiencia, simplificación y automatización (Yuval Noah, 2024).

Desde una perspectiva humanista, la cibercultura no debe ser entendida únicamente como un entorno técnico, que genera oportunidades y desafíos, o que es una herramienta por controlar, ella adquiere otro valor, en cuanto da sentido a la existencia y propone respuestas a problemáticas existenciales, por último, es el denominado nuevo continente al que se está obligado habitar (Spadaro, 2014).

Se vuelve urgente recuperar el valor de la experiencia, la empatía y la diversidad en los entornos digitales, un caso a considerar es, el auge de plataformas colaborativas como Wikipedia representa una manifestación concreta de una cultura con impronta colaborativa y humanista: miles de personas contribuyen voluntariamente al conocimiento colectivo, desafiando jerarquías tradicionales y promoviendo una inteligencia distribuida. Este tipo de prácticas revela que, más allá de los riesgos de automatización, la tecnología puede ser también un vehículo para la cooperación, la creatividad y el pensamiento crítico.

Límites de las categorías tradicionales

En cierto sentido estamos usando moldes preexistentes para realidades totalmente nuevas, lo que redundará en reduccionismos o caricaturas poco precisas, cayendo a veces en discursos moralistas, catastróficos o ingenuos, que no logran captar la complejidad del fenómeno tecnológico contemporáneo. Sabio es el evangelio de Marcos al señalar que: “nadie echa

vino nuevo en odres viejos, porque el vino nuevo rompería los odres, se derramaría y los odres se perderían”. Para nada se debe hablar con certezas sobre las nuevas tecnologías, usando las categorías de siempre. Parece legítimo afirmar que estamos aún en un proceso de comprensión de algo nuevo, aunque ya cuente con una trayectoria significativa en la cultura moderna.

La amplitud del alcance de esta nueva cultura obliga a repensar categorías clásicas como: público, privado, justicia. ¿Qué significa “público” cuando un tuit puede alcanzar a millones en segundos? ¿Cómo redefinimos la “privacidad” en un entorno donde los dispositivos rastrean cada paso? Cada nuevo algoritmo o aplicación abre un espacio de reflexión interdisciplinar que, más que resolver dudas, suele multiplicarlas.

La tecnología no puede ser comprendida únicamente bajo la categoría de herramienta funcional; es también un entorno simbólico que transforma nuestras formas de pensar, sentir y vincularnos. Es así como, el uso de inteligencia artificial en procesos judiciales ha generado controversias éticas que no pueden resolverse apelando únicamente a nociones clásicas de “imparcialidad” o “objetividad”. Algoritmos que predicen reincidencia o asignan puntuaciones de riesgo penal desafían nuestras concepciones tradicionales de justicia, al introducir variables estadísticas en decisiones que afectan vidas humanas. Este tipo de casos exige nuevas categorías que integren la dimensión técnica, social y ética de los sistemas digitales.

Vivimos aún anclados en moldes conceptuales creados para realidades pre-internet: la oposición sujeto-objeto; lo natural frente a lo artificial; la causalidad lineal. Al aplicar estos esquemas a espacios híbridos donde un perfil virtual coexiste con identidades corporales y simbólicas, corremos el riesgo de producir análisis reduccionistas. Aparecen entonces caricaturas de la tecnología como mal autónomo o de la sociedad como víctima pasiva, sin agencia ni capacidad de respuesta.

Este desfase conceptual se traduce en discursos moralistas que presagian apocalipsis tecnocráticos o en apologías ingenuas que celebran cualquier avance sin preguntas críticas. Se requiere en cambio, marcos

analíticos que capten la coevolución entre artefactos digitales y tejido social, reconociendo la agencia distribuida entre diseñadores, usuarios, algoritmos e infraestructuras. O se puede caer en un discurso moralista del uso y desuso de lo virtual y de todas las nuevas tecnologías, debido a ese algo que aún no logra ser comprendido del todo, y pretendiendo orientarlo con las categorías de siempre.

Repensar lo natural y lo artificial

Lo anterior lleva a repensar categorías fundamentales como pueden ser qué es lo natural y artificial, eso permitirá abrirse a aceptar que todo producto de la inteligencia humana pertenece a lo natural, por lo tanto, nos obliga a disolver la barrera entre lo dado por la biología y lo diseñado por el ingenio. Yuval Noah Harari nos recuerda que, si el hombre puede concebir algoritmos de aprendizaje automático, estas creaciones son tan orgánicas como un lenguaje o una herramienta prehistórica. La distinción clásica pierde sentido cuando el ser humano y la máquina dialogan en tiempo real (Yuval Noah, 2016).

En un mismo sentido lo afirmaba Pierre Lévy, al cuestionar la concepción balística que se hacía respecto a las tecnologías y a sus supuestos “impactos” en la cultura y la sociedad, atribuyendo a ambas (tecnología y sociedad) propiedades que no les son propias: A la tecnología un principio activo y autónomo (hasta el momento no lo son) y a la cultura otro pasivo. ¿Es la tecnología un actor autónomo, separado de la sociedad o de la cultura? ¿Es la cultura una entidad pasiva afectada por un agente externo, la técnica? (Levy, 2007).

Esta concepción balística de la que habla Levy, y que en cierto modo intuye Yuval Noah, es una interesante visión que se debería considerar en la reflexión que se pueda hacer de las nuevas tecnologías y de la inteligencia de las máquinas o la denominada Inteligencia Artificial. Más que entidades separadas, tecnología y cultura coevolucionan, se moldean mutuamente y configuran nuevos estilos de vida.

El debate y desfase conceptual entre lo nativo y lo artificial se manifiesta, un buen ejemplo a considerar, en las discusiones sobre neurotecnología y prótesis inteligentes. ¿Dónde termina el cuerpo humano y dónde comienza la máquina cuando un implante cerebral permite controlar dispositivos con el pensamiento? Este tipo de tecnologías desdibujan las fronteras entre lo biológico y lo digital, obligándonos a revisar nociones como autonomía, conciencia o identidad. Persistir en categorías rígidas impide comprender la complejidad de estos fenómenos híbridos y sus implicancias éticas, sociales y filosóficas.

El reduccionismo del yo digital

En un plano distinto emergen las derivas reduccionistas que el ser humano produce sobre su propia condición, al interactuar con las nuevas tecnologías. Pero el reduccionismo, no es exclusiva de estas, también lo puede hacer la política, la economía o lo religioso entre otros. Sin embargo, el entorno digital potencia ciertas formas de simplificación identitaria que merecen atención.

Críticos de estos impactos nocivos y reduccionistas del yo digital son por ejemplo Eric Sadin que afirma que “estaríamos ante la era del individuo tirano; el advenimiento de una condición civilizatoria inédita que muestra la abolición progresiva de todo cimiento común” (Sadin, 2022). Las nuevas tecnologías permiten oportunidades de expansión del conocimiento, de la eficiencia del uso del tiempo, pero a su vez atrofia otras capacidades o características del ser, como pueden ser la de tener paciencia o de errar.

La constitución de un nuevo ethos cultural se propaga aceleradamente, configurando ritos, hábitos y costumbres impregnados de desconfianza y malestar social. Este proceso rompe con el contrato comunitario y erosiona las instituciones tradicionales, y se conforman nuevas creencias centradas en las posibilidades. En este contexto el filósofo francés Sadin advierte de las “representaciones infladas de uno mismo” (Sadin, 2022), como expresión de un sujeto que se concibe en términos de autoafirmación constante y desvinculación de lo colectivo.

Tal dinámica hunde sus raíces en un modelo de liberalismo capitalista radical, donde las instituciones dejan de ser espacios de confianza compartida y se transforman en plataformas de interacción transaccional (Zuboff, 2021). La desconfianza se instala como práctica cotidiana, y el otro deja de ser un interlocutor para convertirse en un auditor o competidor. En lugar de comunidad, emergen microrrelaciones atomizadas, gobernadas por algoritmos que maximizan el engagement y refuerzan la lógica del rendimiento, consolidando así un orden social marcado por la vigilancia y la competencia permanente.

Este desfase conceptual entre lo individual y lo colectivo se evidencia en la forma en que se construye el yo digital. Las plataformas sociales promueven una identidad performativa, fragmentada y cuantificada, donde el valor de la experiencia se mide en likes, vistas y seguidores. En TikTok, la viralidad de un contenido puede convertir a un usuario en referente cultural en cuestión de horas, sin que medie una trayectoria, contexto o reflexión. Esta lógica de exposición constante redefine el sentido del reconocimiento, la autoestima y la pertenencia, desplazando el diálogo interior por métricas externas de validación, y de su propio sentido social o comunitaria de su experiencia.

Dimensión emocional y espiritual en la cibercultura

La comunicación globalizada de carácter virtual promete conexión, pero a menudo refuerza la soledad y la uniformidad afectiva (Han, 2020), es decir la conexión, no garantiza el “encuentro”, entre individuos o comunidades. A nivel emocional y espiritual la cibercultura promueve vínculos superficiales, propicia la conexión, pero no la comunión en la diversidad, todo tiende a ser más monótono, unilateral, uniforme y masivo, no hay cabida para lo diferente, lo distinto. hay una predominancia del presente, el aquí y ahora es lo único que importa mientras que el futuro, la espera y la memoria pierden relevancia.

Lo anterior tiene implicancias en el quehacer político y religioso (espiritual), cuando las relaciones digitales se reducen a flujos de información cuantificables, “me gusta”, “comentarios”, “visualizaciones”, se erosiona el

suelo afectivo donde germinan la empatía, la solidaridad y la experiencia trascendente (Han, 2022).

Eric Sadin afirma al respecto que la psicología individual y colectiva se han visto afectadas por estas nuevas tecnologías de modo que ha ido reconfigurando todas las dimensiones de la existencia:

Hasta qué punto las nuevas tecnologías modificaron nuestras mentalidades casi de modo insidioso, hasta qué punto contribuyeron a la adopción de posiciones inéditas, redefinieron la relación habitual con lo real, con los demás y con gran número de marcos que determinaban hasta ahora la vida en común.... (Sadin, 2022)

Dichas tecnologías ofrecen un poder sin medidas a una humanidad que no sabe bien qué hacer con aquello, ofreciendo responder a las aspiraciones más profundas de inmortalidad, poder y felicidad fáctica (Yuval Noah, 2016). Pero también afectando significativamente la experiencia de vulnerabilidad, de fracaso y de monotonía, tan propias de la existencia, en las cuales se juega la paradoja de la condición humana: su debilidad y su grandeza.

Este panorama reduccionista del yo digital exige, con urgencia, la recuperación de espacios de desconexión respecto de las realidades virtuales y digitales, así como la creación de encuentros presenciales capaces de restablecer los lazos sociales, que restituyen confianzas, revalorizan la diversidad afectiva. Solo mediante esta reapropiación de la experiencia compartida será posible restaurar la hondura de la vivencia emocional y espiritual, reconociendo en la vulnerabilidad, el silencio, la misma desconexión digital, no una carencia, sino la puerta de acceso privilegiada a la alteridad y a la trascendencia.

Un sistema híbrido de prácticas culturales

La cibercultura no es sólo un agregado de aplicaciones o dispositivos, debe concebirse como un sistema integrado de producción, circulación y reproducción cultural. En él confluyen narrativas, símbolos y acciones que se reinventan constantemente en íntima dependencia con infraestructuras

técnicas y lógicas algorítmicas. Una publicación en una red social, puede generar movimientos sociales, viralizar campañas de salud o reformular prácticas académicas; a su vez, las instituciones: gobiernos, medios de comunicación social, universidades, alimentan esa marea digital con discursos normativos y marcos regulatorios.

Comprender esta trama híbrida exige desplegar métodos que combinen análisis de redes sociales, etnografía digital y teoría crítica. Cada publicación, “me gusta” o comentario participa en procesos complejos de validación, visibilidad y exclusión. Ese diálogo entre prácticas culturales y estructuras técnicas da lugar a nuevas formas de subjetividad y de comunidad. en las que lo individual y lo colectivo, lo material y lo simbólico, lo humano y lo no humano, se entrelazan sin jerarquías preestablecidas.

Este sistema híbrido también desafía nuestras categorías tradicionales de autoridad, autenticidad y pertenencia. Otro caso para considerar son los influencers digitales, que operan como nódulos de legitimación cultural, desplazando figuras clásicas como el intelectual público o el líder comunitario. Estos nuevos líderes digitales poseen la capacidad de generar sentido colectivo que no proviene de credenciales institucionales, sino de métricas algorítmicas como el engagement, la viralidad o la fidelización de audiencias. Esta transformación obliga a repensar cómo se construye el prestigio simbólico en la era digital y qué tipo de vínculos sociales emergen de estas nuevas formas de mediación.

En última instancia, repensar la cibercultura como fenómeno humanista implica situar al ser humano en el centro de esta red compleja. No se trata de celebrar acríticamente la innovación tecnológica, sino de reconocer sus riesgos, ambigüedades y potencialidades. La tarea urgente es construir marcos analíticos que permitan comprender cómo se configuran nuevas subjetividades, cómo se rearticulan las comunidades y cómo se redefine el sentido de lo común en un entorno cada vez más mediado por dispositivos inteligentes y plataformas digitales.

La religión y su aporte antropológico a la cibercultura

El corazón se ha vuelto sordo debido al ruido constante de la comunicación e información. Ya no escucha la voz del silencio

Byung Chul Han

En el contexto actual de desafiliación institucional, los sistemas de creencias -tanto personales como colectivos- se ven profundamente influenciados por las nuevas tecnologías, especialmente las de comunicación. Si bien se suele atribuir esta crisis religiosa al descrédito de las estructuras normativas, los estilos de liderazgo o la pérdida de la fe, existe una dimensión más profunda que merece atención: la transformación de las categorías de tiempo y espacio (Han, 2023).

Estas categorías, que han sido constitutivas de la organización humana desde sus orígenes, están siendo reconfiguradas por la lógica digital, que fragmenta la experiencia y diluye las fronteras entre lo público y lo privado, entre el trabajo y el descanso, entre el estar y el no estar (Han, 2022).

Las nuevas tecnologías rompen con esta dualidad complementaria y fragmenta al ser humano en un aparente fluir continuo, donde no se distingue un tiempo y un espacio determinado. Incluso hay un aparente prescindir de ellos (espacio y tiempo), se puede estar y no a la vez, en uno o más momentos en forma simultánea (Han, 2023). Esta disolución de coordenadas existenciales afecta directamente la vivencia religiosa, que históricamente ha ofrecido ritmos, espacios y rituales que estructuran la experiencia humana.

Las tecnologías digitales, especialmente aquellas que median nuestras relaciones, emociones y decisiones, configuran una especie de seudo religión (Carbajal, 2025). En ella se establecen ritos (como los algoritmos de recomendación), sistemas morales (basados en métricas de popularidad) y promesas de salvación (eficiencia, inmortalidad, plenitud). El espacio sagrado es la virtualidad y el canal de comunión es la conectividad. Sin embargo, esta analogía resulta insuficiente desde una perspectiva antropológica rigurosa,

pues carece de los elementos esenciales que definen lo religioso: el principio de la caridad, la gratuidad radical, encuentro con el otro (prójimo), misterio, vínculo con lo trascendente y la experiencia del límite, por nombrar algunos componentes fundamentales.

La religión supone un ritmo y un espacio distinto del que propone una pseudo religión digital, marcada por la producción y el consumo, que propician hiperactividad y la denominada multitasking, síndrome del pensamiento acelerado (Han, 2023). Ella, rescata del tiempo acelerado e indeterminado, e invita al silencio y la contemplación. Conecta lo humano con su origen territorial, es decir lo reconecta con la tierra, el polvo desde donde procede y a donde volverá.

Mutuas contribuciones entre Religión y Nuevas Tecnologías

La pregunta que surge es cuáles son por una parte las interpelaciones (tensiones) que estas nuevas tecnologías suscitan en el ámbito religioso, en especial a las religiones propiamente tal, y cuáles pueden ser los aportes recíprocos entre lo religioso y las nuevas tecnologías de la comunicación.

La religión y la experiencia religiosa apuntan a una relación, una reconexión entre lo humano y lo sagrado, donde se superan los dualismos epistemológicos entre uno y otro ámbito, en dicha relación hay mucho de intuición, de duda, de incertidumbre que conlleva creer y confiar, lo anterior a su vez abre a la experiencia del asombro y la contemplación. Volver a lo religioso, es una invitación a la observación y a escuchar. Para escuchar con el corazón se requiere de la inactividad, el yo que escucha sale de sí mismo para sumergirse en el todo del infinito (Han, 2023).

El corazón, en la tradición bíblica, no se limita a ser el órgano físico, sino que representa la totalidad del ser humano interior, el lugar donde se encuentran la voluntad, el entendimiento y las emociones. Como señala Karl Rahner, “el corazón es el lugar de la experiencia de Dios, donde se produce el encuentro entre la gracia divina y la libertad humana” (Siliézar, 2008). El corazón expresa y acoge la dimensión afectiva de la fe, y se constituye, a su vez, en un espacio o dimensión de resistencia frente a aquello que

deshumaniza y violenta, un refugio donde la dignidad humana se preserva y se fortalece.

Interpelaciones tecnológicas a lo religioso

Las tecnologías digitales, en particular las de comunicación ubicua, han transformado radicalmente las condiciones de posibilidad de la experiencia religiosa. La hiperconectividad, la simultaneidad de espacios, la aceleración del tiempo y la lógica algorítmica desafían las formas tradicionales de ritualidad, contemplación y comunidad. En este nuevo entorno, el sujeto religioso se ve tensionado entre la inmediatez del presente digital y la profundidad del ritmo litúrgico; entre la fragmentación de la atención y la exigencia de escucha interior.

La religión, entendida como práctica simbólica que articula la finitud humana con el deseo de sentido, se ve interpelada por una cultura que tiende a disolver el misterio en datos, a reemplazar la espera por la respuesta instantánea, y a sustituir la gratuidad por la transacción. Esta tensión no implica necesariamente antagonismo, pero sí exige un discernimiento crítico sobre los modos en que lo religioso puede habitar el espacio digital sin perder su densidad antropológica.

Las nuevas tecnologías ponen énfasis en aspectos antropológicos poco considerados por lo religioso, o resignifica otros aspectos, tales como la dimensión comunicacional y cultural, que está presente desde los orígenes de la especie, expresado en dibujos o pintados en cavernas o piedras, hasta la construcción de sistemas sofisticados para comprender los astros o los ciclos lunares y sus efectos en la navegación primitiva.

Yuval Noah Harari, lo expresa muy bien, en su obra *Nexis*, en donde desarrolla precisamente este rasgo cultural y antropológico: Los algoritmos existen desde los tiempos más remotos de las primeras civilizaciones. “Los egipcios usaron algoritmos para medir la inundación y el flujo del río Nilo. Otras civilizaciones utilizaron algoritmos para determinar las estaciones” (Yuval Noah, 2024). La misma Sagrada Escritura, puede ser concebida como

una mediación “tecnológica”, entre Dios, su revelación y destinatarios, quizás las cartas paulinas, son los primeros “email” de la época moderna.

En cierto modo los avances científicos tecnológicos son una expresión de esas necesidades de trascendencia o pseudo trascendencia que el ser humano posee, y aquello expresa otro rasgo antropológico; el de traspasar límites, trascender, que a su manera la cultura tecnológica, especialmente por medio de las redes sociales, la intensifica y los mercados las aprovechan.

La noción de “tecnología con alma” propone una mirada que supera el reduccionismo instrumental y reconoce que los artefactos tecnológicos, al estar insertos en prácticas humanas, pueden vehicular sentido, afectividad y trascendencia. Esta idea se vincula con una teología de la encarnación, donde lo divino se manifiesta en lo humano y lo cotidiano. Si la tecnología es parte del hábitat existencial del ser humano, entonces también puede ser espacio de Revelación, de encuentro con lo sagrado. Esta perspectiva invita a pensar la técnica no como amenaza, sino como posibilidad de mediación espiritual, siempre que se mantenga abierta a la gratuidad, al misterio y al vínculo con el otro.

Aportes de la religión a la cibercultura

Frente a estas transformaciones, las religiones pueden ofrecer contribuciones significativas a la configuración ética y simbólica de la cibercultura. Son diversos los niveles en que ellas pueden aportar. Una de ellas según el investigador de las comunicaciones, Moisés Sbardelotto, será el de ser un agente social mediador ente tecnología, cultura y humanidad, refiriéndose especialmente a la Iglesia Católica, afirmará el investigador:

La Iglesia puede ser uno de esos cuerpos sociales intermedios, fundamentales para condicionar este desarrollo tecnológico. Tal vez no pueda transformarse por completo, porque es una actriz entre varios otros actores sociales que también se cuestionan sobre todo esto. Pero, sin duda, por su historia, su sabiduría, su capilaridad social, la Iglesia tiene un papel muy importante que desempeñar, no sólo

desde el punto de vista de la Santa Sede, sino también en las diversas iglesias locales. (Sbarlodelotto, 2023)

La Iglesia Católica cuenta con una larga tradición en el ámbito de las comunicaciones, y de las agendas valóricas globales, tanto en temas sociales como de la moral de las personas, con aciertos y errores, pero atenta a los signos de los tiempos, y busca diálogos constructivos. Evidencia de ellos son las denominadas Jornadas Mundiales de las Comunicaciones que a solicitud del Papa Paulo VI se realizan desde el año 1967, con una extensa producción de documentos, reflexiones, criterios y trabajo de concientización global.

Otra contribución, en un nivel de mayor profundidad de las religiones, en un sentido teológico propiamente tal, está relacionado con la valoración que hacen del tiempo. El denominado Kairós, categoría teológica, que en la tradición bíblica y patristica se refiere al “momento oportuno” o “momento de gracia” en que Dios irrumpe en la historia, se contrapone al ritmo acelerado.

La cibercultura está marcada por la inmediatez, la multitarea y la constante actualización, en cambio kairós introduce una dimensión alternativa del tiempo: no cronológico, sino cualitativo. Invita a detenerse, discernir y reconocer los momentos significativos que escapan al algoritmo acelerado, y abrir espacios para la elaboración simbólica y la comunión profunda. Este kairós, introduce a su vez otras condiciones para su realización, como son el encontrarse, esperar, contemplar y el silencio. Estos y otros elementos son fundamentales para enriquecer las dimensiones del ser humanos y por lo tanto también del lugar en que se desenvuelve, es decir, la cultura.

Otra contribución de la religión es que re-introduce la experiencia del límite: la vulnerabilidad, la muerte, el misterio. En una cultura digital que tiende a negar la finitud mediante promesas de perfección técnica, inmortalidad virtual y expansión ilimitada, la espiritualidad religiosa recuerda que el sentido no se alcanza por acumulación de datos ni por optimización de procesos, sino por apertura al otro y al trascendente. Como afirma Romano Guardini, “el hombre no se realiza en el dominio, sino en la

aceptación de su límite” (Guardini, 1963). Esta conciencia del límite no es resignación, sino posibilidad de encuentro: El ser humano en cuanto creatura, no puede comprender a su Creador, pero puede abrirse a Él en adoración y confianza, señala Karl Rahner en su teología del misterio.

En este contexto, la religión ofrece una sabiduría que contrasta con la lógica algorítmica, proponiendo una ética del cuidado, del asombro y de la esperanza. Dicha sabiduría, como expresión de plenitud del conocimiento, conecta al humano con una dimensión profunda y muchas veces poco atendida, que es la de su corazón.

El kairós, tiempo de gracia y silencio conecta con el corazón, depositario simbólico de la memoria, las emociones, las valoraciones y las prioridades, "donde está tu tesoro, allí estará también tu corazón" dice el evangelista Mateo. Pero dicho órgano hoy se encuentra sobrecoigido, replegado en sí mismo. Como observa Byung-Chul Han, "si el corazón es el órgano del recuerdo y la memoria, en la era digital estamos absolutamente desprovistos de corazón. Almacenamos cantidades impresionantes de datos e informaciones, pero sin recordar” (Han, 2023). En ese sentido las religiones tienen un aporte que hacer tanto a la humanidad, como a la cultura digital.

Otra contribución de a la cibercultura es que esta ofrece una ética del cuidado y de la hospitalidad. Frente a la lógica de la competencia y la visibilidad, propone el reconocimiento del rostro, la acogida del extraño, la construcción de vínculos no instrumentales. Esta espiritualidad, se constituye en una ética que puede inspirar prácticas tecnológicas más inclusivas, transparentes y orientadas al bien común. Como afirma Emmanuel Levinas, cuyo pensamiento ha influido profundamente en la teología contemporánea: “El rostro del otro en su desnudez y en su miseria es lo que me ordena y me llama” (Levinas, 1982). Esta interpelación ética, que nace del encuentro con el otro, puede trasladarse al diseño de tecnologías que no invisibilicen, que no excluyan, que no cosifiquen.

En este sentido, la hospitalidad religiosa no es sólo una actitud moral, sino una estructura relacional que puede inspirar arquitecturas digitales más humanas. Plataformas que prioricen la escucha, algoritmos que favorezcan la

diversidad, interfaces que promuevan el cuidado mutuo: todos ellos pueden nutrirse de esta tradición espiritual que entiende la acogida como un acto sagrado. La tecnología, entonces, no sería solo herramienta, sino espacio de comunión.

Por último, ofrece una crítica sustantiva a la aparente gratuidad de las plataformas digitales, las cuales, si bien permiten el acceso libre a contenidos, servicios y vínculos, lo hacen mediante un intercambio implícito: el suministro de datos personales, atención e intimidad, posteriormente capitalizados por sistemas de big data (Zuboff, 2021).

La misma autora ha descrito este fenómeno como capitalismo de la vigilancia, en el que “la experiencia humana se convierte en materia prima gratuita para prácticas comerciales ocultas de extracción, predicción y venta” (Zuboff, 2021), declarando así que los usuarios, no son los beneficiarios de servicios, sino más bien una suerte de “materia prima” para los verdaderos beneficiarios de datos, que son las empresas comerciales.

En este modelo, lo que parece gratuito es en realidad una transacción desigual, donde el usuario entrega su subjetividad a cambio de acceso, a su vez se constituye en un medio y no un fin en sí mismo. La religión, en cambio, propone una lógica del don que no se mide en términos de utilidad ni de retorno, sino en apertura, gratuidad y cuidado.

Esta lógica instrumental contrasta con la gratuidad religiosa, entendida como don no condicionado, que no exige retorno ni genera dependencia, sino que promueve comunión y libertad. Aunque la tecnología puede mediar ciertas expresiones religiosas, no sustituye su dimensión encarnada, simbólica y comunitaria, esencial para la experiencia espiritual.

La tesis de este trabajo propone que la religión no solo sobrevive en la era digital, sino que puede ofrecer claves interpretativas profundas para habitar la cibercultura con sentido. La espiritualidad, al recuperar el silencio, la contemplación y la alteridad, puede humanizar los entornos tecnológicos y resistir la lógica del rendimiento y la hiperconexión. La “tecnología con alma” no es una utopía ingenua, sino una invitación a integrar lo técnico con lo ético,

lo simbólico y lo trascendente. En este cruce, se juega no solo el futuro de la religión, sino también el de la humanidad.

Referencias

- Antolinos Asensio, N. (2019). Las nuevas tecnologías y su influencia en las relaciones interpersonales. *Razón Histórica*, (42), 179–193. <https://digitum.um.es/digitum/bitstream/10201/67479/1/LRH%2042.11.pdf>
- Boff, L. (2002). *Tiempo de transcendencia: El ser humano como un proyecto infinito*. Sal Terrae.
- Carbajal, I. (2025). Explorando las fronteras digitales de la religión: Un análisis de las fronteras de la Inteligencia Artificial y la Teología. *Cuadernos de Teología*, 17. <https://doi.org/10.22199/issn.0719-8175-6558>
- Cerda Sanhueza, A. (2020). Acerca de las cosas nuevas: Y las nuevas cosas que interpelan al pensamiento social cristiano. *La Cuestión Social*, 51–76.
- Floridi, L. (2019, 29 de septiembre). *La era de la vida en línea, donde lo real y lo virtual se fusionan* [Entrevista]. J. D'Alessandro, Entrevistador. <https://ihu.unisinos.br/78-noticias/593095-luciano-%EF%AC%82oridi-vou-explicar-a-era-do-onlife-onde-real-e-virtual-se-comfundem>
- Graeber, D., & Wengrow, D. (2022). *El amanecer de todo: Una nueva historia de la humanidad*. Ariel.
- Guardini, R. (1963). *El fin de la modernidad*. Castilla.
- Han, B.-C. (2020). *En el enjambre*. Herder.
- Han, B.-C. (2022). *No cosas*. Penguin Random House.
- Han, B.-C. (2023). *Vida contemplativa*. Taurus.
- Harari, Y. N. (2016). *De animales a dioses: Breve historia de la humanidad*. Debate.
- Harari, Y. N. (2016). *Homo Deus: Breve historia del mañana*. Penguin Random House.
- Harari, Y. N. (2024). *Nexus*. Debate.
- Jaspers, K. (1932). *La filosofía*. Herder.
- Levinas, E. (1982). *Ética e infinito*. Antonio Machado.
- Levy, P. (2007). *Cibercultura*. Anthropos.
- McKinsey Global Institute. (2023). *The future of work in technology*.
- Pérez, F. C. (2023). *Llamados a encontrarnos*. Sal Terrae.
- Sadin, E. (2022). *La era del individuo tirano: El fin del mundo común*. Caja Negra.
- Sbardelotto, M. (2023, 20 de mayo). *Inteligência Artificial e a Igreja: desafios e possibilidades* [Conferencia]. <https://www.ihu.unisinos.br/categorias/159-entrevistas/628554-a-igreja-do-seculo-xxi-e-parte-da-ambiencia-da-inteligencia-artificial-que-gera-transformacoes-sociais-conferencia-de-moises-sbardelotto>

- Siliézar, C. R. (2008). La teología de la gracia en el pensamiento de Karl Rahner. *Cuadernos de Teología*, 137–155.
- Spadaro, A. (2014). *Ciberteología: Pensar el cristianismo en tiempos de red*. Herder.
- Žižek, S. (2015). *Pedir lo imposible*. Yong June Park.
- Zuboff, S. (2021). *La era del capitalismo de la vigilancia*. Paidós.

Pensar a pesar del canto de las cigarras Thinking despite the song of the cicadas

Héctor Inzaurrealde
hector.inzaurrealde@ucu.edu.uy
Universidad Católica del Uruguay
ORCID ID 0009-0007-3420-3067

Resumen: Pensar, en la era de las máquinas, se ha vuelto un acto de resistencia. El presente artículo examina la crisis del pensamiento filosófico en el contexto tecnocrático contemporáneo y su impacto en la educación, la ética y la vida política. A partir de una lectura cruzada del Fedro de Platón —particularmente el mito del canto de las cigarras y la crítica socrática a la escritura—, de las reflexiones de Miguel Pastorino en *Pensar en la era de las máquinas* (2025), y del análisis de Fabio Morandín sobre John Searle y el “cuarto chino”, se propone una reflexión sobre la necesidad de recuperar el pensamiento como praxis emancipadora frente a la automatización de la conciencia. En tiempos de inteligencia artificial y de imperio de los datos, la filosofía se revela como una forma de insurrección cognitiva: un modo de salvaguardar la dignidad humana mediante el juicio, la lentitud y la profundidad. Lejos de ser un lujo académico, pensar se convierte en el núcleo ético-político de la libertad. El texto sostiene que la educación filosófica es hoy una responsabilidad pública, llamada a restaurar la capacidad de discernimiento, el diálogo y la búsqueda de sentido, pilares de toda democracia viva.

Palabras clave: Pensamiento; inteligencia artificial; técnica; educación; filosofía política.

Abstract: Thinking, in the age of machines, has become an act of resistance. This article examines the crisis of philosophical thought within today's technocratic context and its impact on education, ethics, and political life. Drawing on Plato's *Phaedrus*—particularly the myth of the singing cicadas and Socrates' critique of writing—together with Miguel Pastorino's *Pensar en la era de las máquinas* (2025) and Fabio Morandín's analysis of John Searle's "Chinese Room," the paper argues for the need to recover thinking as an emancipatory praxis in the face of the automation of consciousness. In a time dominated by artificial intelligence and the empire of data, philosophy emerges as a form of cognitive insurrection: a way of safeguarding human dignity through judgment, slowness, and depth. Far from being an academic luxury, thinking becomes the ethical and political core of freedom. The text maintains that philosophical education is now a public responsibility aimed at restoring discernment, dialogue, and the search for meaning—foundations of any living democracy.

Keywords: Thinking; artificial intelligence; technology; education; political philosophy.

Introducción

Vivimos tiempos paradójicos. Nunca la humanidad dispuso de tantos datos ni de tantas herramientas cognitivas, pero pocas veces ejerció tan poco el pensamiento crítico. El desarrollo de la inteligencia artificial y la cultura tecnocrática han instalado un nuevo paradigma epistémico: el del cálculo como medida de lo real. Pastorino (2025) advierte que “la creciente falta de pensamiento reside en un proceso que consume la médula misma del hombre contemporáneo: su huida ante el pensar” (p. 4). El sujeto moderno, cautivado por la inmediatez, confunde la velocidad con la comprensión, el procesamiento con el juicio, y la información con la sabiduría.

Pensar, en este escenario, se ha vuelto una forma de disidencia. En medio del ruido global de pantallas, métricas y flujos digitales, pensar significa detenerse. Significa, como Sócrates frente a Fedro, resistir el canto seductor

de las cigarras que invitan al sueño intelectual. El mito, relatado en el *Fedro*, muestra a unos hombres que, fascinados por el canto, olvidaron alimentarse y murieron; transformados en cigarras, dedican su existencia a cantar sin cesar. Sócrates advierte a Fedro que no debe dejarse arrullar por su melodía, pues las cigarras representan el peligro del pensamiento superficial, de la palabra vacía que reemplaza la verdad por el placer de oírse hablar.

Este mito es hoy una metáfora política. El “canto” del que habla Platón encuentra su eco contemporáneo en la fascinación por las tecnologías de la información. La humanidad, embelesada por la promesa del control total de los datos y por la inmediatez de las respuestas automáticas, corre el riesgo de adormecer su conciencia crítica. Pensar a pesar del canto de las cigarras es, entonces, un acto de resistencia frente al hechizo tecnocrático: un llamado a mantener despierta la razón cuando todo conspira para anestésicarla.

En este sentido, la introducción delimita ya una tesis fundamental: no se trata de rechazar la técnica en bloque, sino de advertir que su contexto cultural puede favorecer el pensamiento superficial si no hay un ejercicio consciente de vigilancia intelectual. Es justamente esta tensión —entre fascinación tecnológica y responsabilidad reflexiva— la que guiará el desarrollo del artículo.

El canto de las cigarras y el riesgo del sueño intelectual

Platón ubica en el *Fedro* una de las primeras críticas filosóficas a la técnica: la desconfianza hacia la escritura como sustituto del pensamiento vivo. Sócrates le dice a Fedro que la escritura “parece dotada de inteligencia, pero si se le pregunta algo, guarda silencio” (*Fedro*, 275d). La escritura no puede defenderse ni explicar, y quien confía en ella “dejará de ejercitar su memoria y se volverá olvidadizo” (*Fedro*, 275a).

Salina (2024) advierte que el filósofo griego no condena la escritura, sino su uso acrítico. Lo que teme es el desplazamiento del logos vivo —el diálogo del alma consigo misma— por una técnica que simula pensamiento sin comprender. Este es, también, el dilema contemporáneo de la inteligencia

artificial: “una memoria sin interioridad, un lenguaje sin sujeto” (Salina, 2024, p. 15).

Derrida reinterpreta esta tensión con el término *pharmakon*: la técnica es remedio y veneno a la vez. Puede conservar la memoria, pero también anular su ejercicio. En la IA, esa ambivalencia se radicaliza: su potencia para ampliar la cognición humana convive con la tentación de reemplazarla. En palabras de Pastorino (2025), “la inteligencia artificial promete liberarnos del esfuerzo de pensar, pero corre el riesgo de liberarnos del pensamiento mismo” (p. 7).

El mito de las cigarras encarna, así, la advertencia sobre el placer del conocimiento sin esfuerzo. Las cigarras fueron, según Platón, humanos que se olvidaron de vivir mientras cantaban; los hombres contemporáneos, encantados por el ruido digital, corren el riesgo de olvidarse de pensar mientras opinan.

Desde un punto de vista político, el canto de las cigarras representa el consenso del entretenimiento: un pueblo adormecido por el flujo incesante de información. La tarea del filósofo, como Sócrates, es mantener la vigilia y devolver a la palabra su carácter de búsqueda.

El pensamiento, diría Adela Cortina (2013), “no consiste en repetir ideas, sino en alumbrar razones para la acción moral” (p. 45). Pensar no es un acto contemplativo de evasión, sino una forma activa de responsabilidad frente al otro. Dejarse arrullar por el canto de las máquinas es abdicar de esa responsabilidad.

Por tanto, el riesgo no está en la tecnología misma, sino en la reducción del pensamiento a mera repetición o consumo informacional. Las cigarras simbolizan una actitud humana: la preferencia por el sonido sobre el sentido, por la inmediatez sobre la comprensión. Por eso, pensar implica un esfuerzo deliberado de atención y diálogo, es decir, una práctica que no puede quedar sustituida por ningún dispositivo técnico.

El imperio de los datos y la pérdida del juicio reflexivo

Pastorino (2025) profundiza la crítica heideggeriana al dominio de la técnica: “la tecnocracia no es solo un sistema de gestión, es una ideología del control que sustituye el juicio por el cálculo” (p. 18). En la lógica tecnocrática, la realidad se convierte en dato, el conocimiento en rendimiento, y la verdad en eficiencia.

Heidegger (1954/1994) había anticipado este fenómeno al hablar del Gestell —el “enmarcamiento”—: “La esencia de la técnica no es nada técnico. El peligro supremo es que el hombre sea tomado como fondo disponible, como recurso” (p. 23).

Para el pensador alemán, la técnica moderna no es una herramienta neutral, sino una forma de revelar el mundo que transforma todo en objeto útil, incluso al ser humano. Pastorino retoma esta advertencia para describir cómo la inteligencia artificial intensifica el proceso: “El hombre, que creía dominar la técnica, descubre que es ella quien lo domina, modelando su modo de pensar y de vivir” (2025, p. 20).

La pérdida del sentido se produce cuando el Gestell sustituye la experiencia del ser por la gestión de lo disponible. La educación, la política y la cultura se orientan al rendimiento, mientras el alma —como diría Platón— olvida mirar hacia el bien.

Heidegger ofrece, sin embargo, un horizonte alternativo en su concepto de Gelassenheit (serenidad): “Nos vemos entregados a la técnica cuando la consideramos algo neutral, porque esta idea nos ciega frente a su esencia” (Heidegger, 1954, p. 7).

Pastorino interpreta este llamado a la serenidad no como pasividad, sino como actitud crítica: “serenidad no es desconexión del mundo técnico, sino libertad frente a su hechizo” (2025, p. 45). Recuperar la serenidad implica reaprender a usar la técnica sin pertenecerle, discerniendo su sentido humano. Byung-Chul Han (2013) coincide en que la hipertransparencia digital elimina la distancia necesaria para el pensamiento: “la sociedad de la información no

permite el secreto, y sin secreto no hay pensamiento profundo” (p. 33). En un mundo sin silencio, el pensamiento deviene cálculo.

La filosofía, por tanto, debe enseñar a habitar la técnica sin reducirse a ella. Su función política consiste en recordar que no todo lo medible es valioso, ni todo lo eficiente es bueno. En palabras de Nussbaum (2010), “las democracias necesitan de las humanidades porque sin juicio no hay libertad” (p. 27).

En conclusión, no puede sostenerse que toda cultura contemporánea esté dominada por la tecnocracia, pero sí que existe una tendencia estructural hacia la cuantificación y la eficiencia que puede debilitar el juicio reflexivo si no se contrarresta con prácticas deliberadas de pensamiento crítico. Por eso, el desafío no es eliminar la técnica, sino formar ciudadanos capaces de situarla dentro de un horizonte humano de sentido.

Pensar no es procesar información: Searle y el “cuarto chino”

La reflexión de Searle, reinterpretada por Morandín (2023), muestra que el procesamiento de información no equivale a comprensión. En el experimento del “cuarto chino”, una persona que no entiende el idioma sigue un conjunto de reglas que le permite responder correctamente en chino, sin entender el sentido de las palabras. “Searle argumenta que la máquina no piensa; sólo sigue reglas formales. Puede simular comprensión, pero no entender” (Morandín, 2023, p. 195).

Esta distinción entre sintaxis y semántica —entre manipular signos y comprender su significado— es fundamental para delimitar el lugar del pensamiento humano. Las máquinas pueden producir frases, pero no significado; pueden organizar información, pero no habitar el mundo que esa información describe. Heidegger (1955) lo expresó de modo más radical: la esencia de la técnica no reside en su utilidad, sino en su capacidad de revelar el mundo. En la medida en que el hombre olvida esa revelación, se convierte él mismo en objeto técnico.

Morandín (2023) añade que incluso los sistemas de inteligencia artificial más avanzados “no tienen una comprensión original de su existencia o su ser en el mundo” (p. 195). No hay en ellos autoconciencia ni intención. Desde una perspectiva heideggeriana, carecen de Dasein: no son entes que se comprendan a sí mismos.

El riesgo político surge cuando el ser humano imita esa condición. Cuando el ciudadano deja de pensar, cuando repite fórmulas, cuando acepta sin juicio los discursos producidos por la máquina, se convierte él mismo en un sistema sintáctico. El problema no es que la IA carezca de conciencia, sino que nosotros abandonemos la nuestra.

Desde esta óptica, el “cuarto chino” se convierte en una alegoría social. Vivimos encerrados en un sistema de signos que interpretamos sin comprender, guiados por algoritmos que nos devuelven nuestras propias palabras. El pensamiento humano, por el contrario, no es un algoritmo sino una apertura al sentido. Penrose (1994) y Chalmers (1996) sostienen que la conciencia humana no se reduce a un sistema computacional, pues involucra una experiencia subjetiva irreducible: el qualia, el “cómo se siente ser”.

Pastorino (2025) retoma esta distinción: “El pensamiento auténtico no es procesamiento lógico, sino experiencia de sentido; no calcula, comprende” (p. 25). Lo que está en juego no es la capacidad operativa de la IA, sino la reducción del hombre a una máquina entre máquinas. El pensamiento crítico consiste en romper esa clausura, en recuperar la capacidad de interpretar.

Morandín (2023) sugiere que la cuestión no es si las máquinas pueden ser conscientes, sino “qué significa ser en el mundo” (p. 194). Pensar, en su sentido más pleno, es habitar el mundo de manera reflexiva, no funcional. La filosofía, por tanto, no compite con la técnica: la interroga. En lugar de imitar la eficiencia de las máquinas, debe enseñar a vivir humanamente en un mundo técnico. Pensar, entonces, no es una función biológica, sino un acto ético. Implica responder: no ante una base de datos, sino ante el rostro del otro.

Por lo anteriormente dicho es que, afirmo que la amenaza no consiste en que la IA produzca información, sino en que los seres humanos abandonen la tarea de comprender. La filosofía aparece, así, como un ejercicio indispensable de interpretación, juicio y responsabilidad. Pensar es lo que permite evitar que la vida pública derive en mera repetición automática de discursos prefabricados y en manipulaciones algorítmicas.

Ética, educación y resistencia del pensamiento

Si la técnica tiende a automatizar, la educación debe humanizar. Pastorino (2025) sostiene que “la filosofía no se estudia para tener una opinión, sino para poder distinguir entre lo opinable y lo fundamental” (p. 29). Esa distinción es la base del juicio ético y político. Educar filosóficamente significa formar personas capaces de pensar con rigor y libertad, de sostener el disenso y de resistir la simplificación.

La educación contemporánea enfrenta el riesgo de convertirse en mera capacitación técnica. Las políticas educativas, presionadas por los indicadores de desempeño, promueven competencias medibles, pero descuidan la formación del juicio. En este sentido, Pastorino denuncia “una educación que deja de pensarse a sí misma desde una perspectiva sapiencial” (p. 18). La crisis actual de la educación, subordinada a indicadores de rendimiento, refleja el triunfo de la racionalidad instrumental. Paulo Freire (1970/2005) denunció esa tendencia como una “educación bancaria”, donde el conocimiento se deposita en el alumno sin diálogo. Frente a ella, propuso una pedagogía de la palabra: “nadie educa a nadie, nadie se educa solo, los hombres se educan entre sí, mediatizados por el mundo” (p. 81).

La filosofía puede devolver a la educación su vocación de humanidad. En términos éticos, pensar es cuidar: cuidar del sentido, de la verdad y de la existencia del otro. Ortega y Gasset (2014) advertía contra “la barbarie del especialismo” (p. 192), donde el saber se fragmenta y pierde horizonte. El filósofo, en cambio, busca la totalidad; la escuela filosófica no enseña respuestas, sino preguntas. Preguntas, que suscitan la ruptura con el canto romántico de las sirenas e impulsan una actitud crítica y autónoma, frente a la ansiada verdad.

En este punto, la convergencia entre Platón y Pastorino resulta fecunda. El primero enseñaba que el conocimiento sin virtud es peligroso; el segundo, que la técnica sin reflexión es ciega. Ambos coinciden en que la educación no consiste en acumular información, sino en orientar la mirada hacia el bien. Adela Cortina (2022) subraya que “la ética no puede ser reemplazada por algoritmos, porque exige deliberar sobre fines, no solo calcular medios” (p. 67). Enseñar a pensar éticamente implica enseñar a deliberar.

Nussbaum (2010) y Ordine (2013) coinciden en que las Humanidades son el corazón de una ciudadanía libre. Ordine recordaba que “la utilidad de lo inútil” reside en aquello que no se mide, pero nos humaniza: la poesía, la filosofía, la contemplación. Pastorino (2025) denomina a esta tarea “resistencia sapiencial”: la defensa del pensamiento lento frente a la prisa técnica.

Desde la perspectiva política, la educación filosófica es una forma de resistencia cultural. Frente a la homogeneización del pensamiento impuesta por los algoritmos, educar para el juicio significa preservar la diversidad de la razón. Como sostiene Pastorino, “en tiempos de sedentarismo cognitivo, la filosofía es gimnasia del espíritu” (p. 48). Pensar, entonces, es un ejercicio: un entrenamiento en libertad.

La filosofía no forma consumidores de verdades, sino productores de sentido. Por eso, su papel en la educación no es accesorio, sino fundacional. Si la escuela del futuro no enseña a pensar, se convertirá en un laboratorio de obediencia digital.

La educación filosófica no es únicamente un campo académico, sino un dispositivo público que fortalece la autonomía ciudadana. En un contexto donde la opinión instantánea parece sustituir al juicio fundamentado, la enseñanza del pensamiento crítico se vuelve una herramienta imprescindible para sostener sociedades democráticas capaces de deliberar y no solo de reaccionar.

Conclusiones

Pensar es un acto político. En una sociedad dominada por la automatización, mantener viva la reflexión es ejercer resistencia frente al conformismo. Pastorino (2025) lo formula con claridad: “La filosofía es el único antídoto contra la banalidad del mal y la automatización de la conciencia” (p. 48).

El mal, en clave arendtiana, surge de la ausencia de pensamiento. No es producto del odio, sino de la inercia. El sedentarismo cognitivo que denuncia Pastorino es el mismo que Arendt (2005) atribuye a Eichmann: la incapacidad de pensar desde la perspectiva del otro. Este vínculo se vuelve más explícito en Eichmann en Jerusalén (Arendt, 1963/2003), donde la autora desarrolla la noción de “banalidad del mal” para mostrar que las acciones de Eichmann no provinieron de un odio fanático, sino de una ausencia radical de pensamiento crítico y juicio moral. En este sentido, la tesis arendtiana refuerza la idea de que la desactivación del pensamiento puede habilitar formas de obediencia ciega y burocrática, con consecuencias devastadoras para la vida pública.

El pensamiento filosófico, en cambio, interrumpe la automatización. Donde la técnica ordena, el pensamiento interroga; donde los algoritmos predicen, el juicio evalúa. Esta es la dimensión política del pensar: su capacidad de abrir espacio para la libertad.

El mito de las cigarras recobra, entonces, toda su fuerza simbólica. Pensar “a pesar de su canto” es mantener la vigilia frente a los encantos de la inmediatez y del rendimiento. El filósofo, como Ulises, debe atarse al mástil del diálogo para no sucumbir al canto de las sirenas tecnológicas.

En la era de las máquinas, la filosofía no es un residuo del pasado, sino una anticipación del futuro. Allí donde los datos gobiernan, la filosofía recuerda la primacía del sentido; donde el mundo se mide en algoritmos, el pensamiento restaura el lenguaje del alma.

Educar para pensar es, por tanto, el acto político más radical. Supone defender la autonomía del espíritu frente a la programación de la mente.

Significa enseñar que el pensamiento no se delega ni se terceriza. Que pensar, en su raíz más honda, es cuidar de lo humano.

El pensamiento filosófico es, en última instancia, una forma de libertad. Pastorino (2025) recoge de Heidegger la noción de *Gelassenheit* — la serenidad como actitud crítica frente a la técnica— y la redefine como un “estado de atención lúcida” (p. 47). Serenidad no es desconexión, sino discernimiento; no es nostalgia del pasado, sino claridad ante el presente.

Heidegger (1955) advirtió: “Estamos más a merced de la técnica cuando la consideramos algo neutral, porque esta idea nos ciega frente a su esencia” (p. 7). El pensador no llama a rechazar la técnica, sino a humanizarla, a comprender que su poder no radica en la eficiencia, sino en su capacidad de moldear nuestra relación con el ser.

Pensar a pesar del canto de las cigarras es pensar sin adormecerse ante el ruido del progreso. Es mantener viva la llama del diálogo en medio del zumbido de las máquinas.

En primer lugar, “pensar a pesar del canto de las cigarras” significa mantener la vigilia interior frente a la distracción seductora del mundo contemporáneo. Así como en el *Fedro* (258c-259d) las cigarras encarnan la tentación de perder la conciencia del vivir mientras se oye su canto, hoy simbolizan la fascinación por la rapidez, la reacción inmediata y la superficialidad informacional. Pensar implica interrumpir ese automatismo, ejercitar la demora y sostener la pregunta cuando todo invita a la respuesta instantánea. En este sentido, pensar no es un acto espontáneo, sino una práctica deliberada y esforzada.

En segundo lugar, la posición que este artículo defiende es que el pensamiento filosófico constituye una práctica emancipadora frente a la automatización de la conciencia. No se trata de oponer técnica y humanidad como realidades incompatibles, sino de afirmar que el pensamiento —como juicio, interpretación y responsabilidad— no puede ser delegado a ningún dispositivo. La filosofía preserva la libertad porque impide que la conciencia se vuelva pasiva o programada. En un mundo donde la técnica ofrece

soluciones inmediatas, el pensamiento recuerda que comprender es más valioso que simplemente procesar.

En tercer lugar, “pensar a pesar del canto de las cigarras” significa también asumir la libertad de profundizar frente a la tendencia cultural a la superficialidad, ejercitar la capacidad de evaluar, discriminar y deliberar. Es un gesto ético —porque implica responder ante el otro— y político —porque abre la posibilidad del disenso y la vida democrática—. El pensamiento aparece, así, no como un lujo académico, sino como un acto ciudadano y una forma de responsabilidad pública.

Desde esta perspectiva, en el ámbito educativo, este argumento exige recuperar el desarrollo del juicio como objetivo pedagógico central. La formación filosófica —en escuelas, liceos y universidades— debe cultivar la capacidad de interpretar, argumentar y dialogar. Esto implica integrar espacios de deliberación ética, análisis crítico y discusión reflexiva, evitando que la educación se reduzca a mera capacitación técnica orientada únicamente al rendimiento o la eficiencia o a la tendencia a la memorización de contenidos. Pensar es aprender a preguntar antes que a repetir. Aprender es desear el saber, no la utilidad.

Asimismo, para la vida pública, una ciudadanía capaz de pensar críticamente fortalece la democracia. Frente a la circulación instantánea de información, la polarización y la manipulación algorítmica, el pensamiento reflexivo permite matizar opiniones, evaluar fuentes y sostener el debate democrático. Pensar no es un ejercicio privado, sino una condición para la convivencia social y para la participación responsable en la esfera pública.

Finalmente, en la relación con la técnica, este enfoque propone una cultura de uso responsable: la inteligencia artificial puede ser una herramienta poderosa siempre que no sustituya la reflexión humana. La serenidad —entendida como atención lúcida— permite beneficiarse de la técnica sin quedar subordinados a ella. De este modo, pensar se convierte en la clave para habitar un mundo tecnológico de manera verdaderamente humana, preservando la dignidad, la libertad y el sentido.

En palabras de Han (2013), “solo quien sabe demorarse puede pensar verdaderamente” (p. 29). Pensar a pesar del canto de las cigarras es pensar por sí mismo. Y ese gesto, hoy, es revolucionario.

Referencias

- Arendt, H. (2003). *Eichmann en Jerusalén: Un estudio sobre la banalidad del mal* (C. Ribalta, Trad.). Lumen. (Obra original publicada en 1963).
- Arendt, H. (2005). *La condición humana*. Paidós.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Cortina, A. (2013). *¿Para qué sirve realmente la ética?* Paidós.
- Derrida, J. (1972). *La farmacia de Platón* (C. Ceretti, Trad.). Siglo XXI.
- Freire, P. (2005). *Pedagogía del oprimido*. Siglo XXI. (Obra original publicada en 1970).
- Han, B.-C. (2013). *La sociedad de la transparencia*. Herder.
- Heidegger, M. (1955). *Serenidad*. Herder.
- Heidegger, M. (1994). La pregunta por la técnica. En *Conferencias y artículos*. Herder. (Obra original publicada en 1954).
- Morandín-Ahuerma, F. (2023). Conciencia e inteligencia artificial: Heidegger, Searle y Bostrom. *Stoa*, 14(28), 189–209. <https://doi.org/10.25009/st.2023.28.2765>
- Nussbaum, M. C. (2010). *Sin fines de lucro: Por qué la democracia necesita de las humanidades*. Katz Editores.
- Ordine, N. (2013). *La utilidad de lo inútil*. Acantilado.
- Ortega y Gasset, J. (2014). *Meditación de la técnica y otros ensayos*. Alianza Editorial.
- Pastorino, M. (2025). *Pensar en la era de las máquinas*. Taurus.
- Penrose, R. (1994). *Shadows of the mind: A search for the missing science of consciousness*. Oxford University Press.
- Platón. (1988). Fedro. En *Diálogos* (Vol. III, C. García Gual, Trad.). Gredos.
- Salina, C. (2024). *Lo que Platón tiene para decirnos sobre la Inteligencia Artificial*. CFE/DGES-ANEP.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/S0140525X00005756>
- Vernant, J.-P. (1992). *Los orígenes del pensamiento griego*. Fondo de Cultura Económica.

Inteligencia artificial y creatividad computacional: El rol del artista en el escenario poshumanista

Artificial intelligence and computational creativity: The role of the artist in the posthumanist era

Iván Vanioff

ivan.vanioff@gmail.com

Universidad Nacional del Nordeste

ORCID ID 0000-0001-7128-4344

Resumen: El artículo analiza la reconfiguración ontológica y práctica que atraviesa la creación artística ante la irrupción de la inteligencia artificial. La tesis central sostiene que la integración de sistemas algorítmicos en el arte no es meramente instrumental, sino que desplaza la figura del artista desde concepciones clásicas (inspiración divina, genio excepcional y expresión subjetiva) hacia un nuevo rol de diseñador de ecologías generativas. En este escenario poshumanista, la autoría deja de ser una propiedad exclusiva del sujeto humano para convertirse en una competencia distribuida, híbrida y colaborativa, donde el artista actúa como mediador y curador de procesos estocásticos.

El argumento se desarrolla a través de un recorrido histórico que comienza diagnosticando los mitos fundacionales de la creatividad artística, los cuales situaban al arte fuera del alcance de la razón. Se contrasta esta visión con la secularización impulsada posteriormente por la psicología y las neurociencias, disciplinas que permitieron objetivar la creatividad entendiéndola como un proceso cognitivo complejo, modelable y medible.

Sobre esta base científica, el texto incorpora modelos teóricos de las ciencias cognitivas para distinguir tres formas de creatividad: combinacional, exploratoria y transformacional. Se argumenta que, gracias al avance del aprendizaje automático y las redes neuronales profundas, la inteligencia artificial actual ha superado la rígida

ejecución de reglas predefinidas. Los sistemas contemporáneos poseen ahora una autonomía técnica que les permite generar novedad mediante la combinación y la exploración de espacios latentes. Sin embargo, se advierte que las máquinas aún carecen de la intencionalidad necesaria para alcanzar una creatividad transformacional plena, aquella capaz de alterar los paradigmas culturales de sentido.

La discusión culmina analizando cómo este panorama técnico redefine la práctica artística. Se plantea que el arte transita de la producción y ejecución de la obra hacia una lógica de selección y curaduría. En lo que se denomina una "estética artificial", el valor creativo ya no reside solo en la acción del artista, sino en su capacidad de generar y seleccionar resultados algorítmicos. Asimismo, surge una co-creación donde la autenticidad de la obra emerge del diálogo situado entre la instrucción humana y la interpretación de la máquina.

En conclusión, el trabajo postula que el artista poshumano abandona la ejecución directa para convertirse en un arquitecto de decisiones. Esta transformación no vuelve obsoleto al ser humano, sino que lo obliga a redefinir su identidad creativa no como una esencia aislada, sino como una potencia que evoluciona y se expande en simbiosis con la técnica.

Palabras clave: Transhumanismo; humanismo; creación artística; arte contemporáneo; cibernético.

Abstract: This article analyzes the ontological and practical reconfiguration that artistic creation is undergoing due to the advent of artificial intelligence. The central thesis argues that the integration of algorithmic systems into art is not merely instrumental but rather shifts the figure of the artist from classical conceptions (divine inspiration, exceptional genius, and subjective expression) toward a new role as designer of generative ecologies. In this posthumanist scenario, authorship ceases to be the exclusive domain of the human subject, becoming instead a distributed, hybrid, and collaborative competence, where the artist acts as a mediator and curator of stochastic processes.

The argument unfolds through a historical overview that begins by examining the foundational myths of artistic creativity, which placed art outside the realm of reason. This view is contrasted with the secularization subsequently promoted by psychology and neuroscience, disciplines that enabled the objectification of creativity by understanding it as a complex, modelable, and measurable cognitive process.

Building on this scientific foundation, the text incorporates theoretical models from cognitive science to distinguish three forms of creativity: combinational, exploratory, and transformational. It argues that, thanks to advances in machine learning and deep neural networks, current artificial intelligence has transcended the rigid execution of predefined rules. Contemporary systems now possess a technical autonomy that allows them to generate novelty through the combination and exploration of latent spaces. However, it is observed that machines still lack the intentionality necessary to achieve full transformational creativity, capable of altering cultural paradigms of meaning.

The discussion concludes by analyzing how this technical landscape redefines artistic practice. It posits that art is shifting from the production and execution of the work toward a logic of selection and curation. In what is termed "artificial aesthetics," creative value no longer resides solely in the artist's action, but rather in their capacity to generate and select algorithmic results. Furthermore, a co-creation emerges where the authenticity of the work arises from the situated dialogue between human instruction and the machine's interpretation.

In conclusion, the article posits that the posthuman artist abandons direct execution to become an architect of decisions. This transformation does not render humanity obsolete but rather compels us to redefine our creative identity not as an isolated essence, but as a power that evolves and expands in symbiosis with technology.

Keywords: Transhumanism; humanism; artistic creation; contemporary art; cyborg.

Introducción¹

La relación entre arte, creatividad e inteligencia artificial constituye uno de los debates más complejos y urgentes del pensamiento contemporáneo. A lo

¹ Este artículo fue elaborado con asistencia de herramientas de inteligencia artificial (IA). Toda la información presentada ha sido obtenida exclusivamente de las fuentes mencionadas y debidamente referenciadas. El uso de la IA en esta investigación cumple con las normativas y protocolos internacionales para el uso ético y responsable de la IA en el ámbito académico y científico, tales como las directrices de la UNESCO para la Ética de la Inteligencia Artificial (2021), los Principios de IA Responsable de la OCDE (2019), y las recomendaciones del

largo de los siglos, el arte ha sido interpretado como un fenómeno que trasciende la razón, vinculado a la inspiración divina, al genio individual o a los misterios de la subjetividad. Estas concepciones, arraigadas en la tradición occidental, dieron origen a creencias que situaron la creatividad fuera del ámbito del conocimiento racional. El presente trabajo se propone analizar esta transformación histórica y epistemológica de la creatividad desde una perspectiva interdisciplinaria que articula filosofía, psicología, neurociencias, ciencias cognitivas y teoría del arte.

La hipótesis central de este trabajo postula que la integración de la inteligencia artificial en el arte no supone una mera actualización técnica, sino una reconfiguración ontológica del rol del artista. Lejos de la concepción romántica del genio solitario que plasma su interioridad en un objeto cerrado, el creador contemporáneo deviene un diseñador de ecologías generativas. En este escenario poshumanista, su función principal se desplaza desde la ejecución manual hacia la mediación y la curaduría de procesos algorítmicos.

El artista ya no produce la obra de manera directa y finalista, sino que programa las condiciones de posibilidad (parámetros, datos y reglas de aprendizaje) para que lo estético emerja de la interacción dinámica con la máquina. De este modo, la autoría se redefine como una competencia distribuida y relacional: el artista orquesta un proceso de co-evolución donde la creatividad surge de la negociación constante entre la intención humana y la agencia computacional, transformando el arte en un laboratorio abierto de inteligencias híbridas.

El artículo se estructura en cinco secciones articuladas de manera progresiva. La primera sección analiza las concepciones tradicionales del arte como expresión divina, genio individual o manifestación de la interioridad. Desde Platón hasta el Romanticismo, la creatividad fue entendida como un don sobrenatural o una fuerza irracional. Al revisar estas narrativas se revela

Grupo de Expertos de la Comisión Europea sobre IA Ética y Confiable (2019). Además, se han seguido las prácticas sugeridas por el Committee on Publication Ethics (COPE) para asegurar la transparencia y la integridad del proceso de publicación.

cómo configuraron la idea del artista excepcional y por qué desmitificar la creación es clave para pensarla en la era tecnológica.

En la segunda sección se explorará el tránsito de la creatividad desde las creencias mencionadas en la sección anterior hacia su estudio científico. La psicología la define como proceso cognitivo y motivacional (Guilford, Csikszentmihalyi, Gardner), mientras que las neurociencias la conciben como interacción entre redes neuronales, emoción y cognición. Este enfoque interdisciplinario transforma la creatividad en una capacidad conocible y natural, estableciendo las bases para comprender su funcionamiento y su potencial relación con la inteligencia artificial.

Luego, en la tercera sección, se presentará la teoría de la creatividad computacional de Margaret Boden, que distingue tres formas: combinacional, exploratoria y transformacional. Al integrar ciencias cognitivas, filosofía e inteligencia artificial, Boden propone un modelo que explica cómo la mente humana genera novedad y cómo estos procesos pueden replicarse parcialmente en sistemas artificiales. La autora redefine la creatividad como fenómeno cognitivo contextual y socialmente validado, abriendo el camino hacia una creatividad híbrida humano-máquina.

Seguidamente, la cuarta sección examinará cómo los sistemas de inteligencia artificial, mediante aprendizaje automático y redes neuronales, reproducen procesos creativos humanos. Las máquinas combinan y exploran datos, generando resultados originales que reflejan creatividad combinacional y exploratoria. Sin embargo, aún carecen de creatividad transformacional, aquella que altera los marcos conceptuales de una cultura. Así, la IA amplía las posibilidades expresivas y experimentales sin reemplazar la capacidad humana de producir significados radicalmente nuevos.

Finalmente, en la quinta sección, se analizará cómo la irrupción del arte con IA redefine la figura del artista y el proceso creativo. El creador se convierte en programador y curador de sistemas generativos, compartiendo agencia con la máquina. Desde el poshumanismo (Haraway, Zyglinska, Astobiza), el arte emerge como un proceso colaborativo y relacional donde

lo humano se entiende como ensamblaje técnico, orgánico y cultural, disolviendo los límites entre sujeto, obra y tecnología.

En conjunto, este recorrido permite observar un desplazamiento decisivo: la creatividad deja de entenderse como emanación interior o producto de una mente individual para concebirse como un proceso emergente, distribuido y tecnológicamente mediado. La inteligencia artificial, lejos de anular la dimensión humana del arte, la expande hacia territorios inéditos de experimentación estética y cognitiva. La creatividad, en la era de la inteligencia artificial, se revela así como un fenómeno híbrido que desafía las distinciones entre naturaleza y técnica, entre sujeto y máquina, entre lo humano y lo artificial.

Del don divino al genio romántico: genealogía de los mitos sobre la creatividad

A lo largo de la historia, la creación artística ha sido interpretada como un fenómeno misterioso, vinculado a fuerzas sobrenaturales o a dones excepcionales. Antes de la consolidación de una mirada científica sobre la creatividad, el arte se explicaba mediante relatos míticos que atribuían su origen a la inspiración divina, al genio individual o a la expresión interior del ser humano. Estas concepciones, aunque cargadas de simbolismo, configuraron una tradición que aún condiciona nuestra comprensión del acto creativo. Examinar estas narrativas permite reconocer cómo se construyeron los imaginarios que separaron el arte de la razón, y comprender por qué desmitificar la creación resulta indispensable para pensar la creatividad en la era de la inteligencia artificial.

Respecto a la primera concepción sobre la creación artística, entendida como inspiración divina, Stefan Zweig (2023) sostiene que existe una visión donde la obra se comprende como un milagro sobrehumano, una irrupción de lo eterno en lo efímero. El artista es visto como un mediador entre el mundo sensible y una fuente divina de inspiración, un sujeto que, en un estado de éxtasis, traduce lo inefable en forma estética. Este paradigma, heredero del pensamiento platónico, concibe el arte como un vehículo de revelación espiritual. Así entendida, “la creación artística es un acto

sobrenatural en una esfera espiritual que se sustrae a toda observación” (Zweig, 2023, p. 8).

En la tradición judeocristiana el modelo arquetípico es la creación *ex nihilo* de Dios, que genera el mundo a partir de la nada y establece el paradigma de la creación absoluta. Recordemos que Platón (1988) sostenía que el proceso creativo del poeta no es un acto de habilidad técnica o de conocimiento racional, sino un estado de inspiración divina o posesión. En *Ion* argumenta que el poeta está fuera de sí, actuando como un mero intérprete o canal de las Musas, sin que su propia inteligencia o sabiduría intervenga.

Margaret Boden (2004) denomina a esta interpretación “inspiracionismo”, pues presupone que la creatividad proviene de una fuerza externa e inexplicable, más cercana al milagro que al pensamiento racional. En su análisis, Boden subraya que esta idea refleja una concepción premoderna del acto creador, en la cual el artista no controla el proceso creativo, sino que lo experimenta como una forma de posesión o iluminación. Este enfoque sugiere que la obra no surge de un aprendizaje o de la práctica, sino de un momento de revelación súbita que el creador apenas logra traducir en materia artística. Así, la figura del artista es pasiva frente al misterio de la creación, un intermediario que canaliza lo divino sin pretender explicarlo ni dominarlo. Este planteamiento prepara el paso hacia una comprensión más humana del poder creador, en la que la inspiración se interioriza en la conciencia individual: la figura del genio.

La segunda concepción mitológica de la creatividad artística remite a la figura del genio. El romanticismo seculariza la noción de lo divino al situar en el genio artístico la fuente de la creación. El genio, según Zweig (2023), es quien supera los límites de la mortalidad mediante la potencia creadora, impulsado por una energía interior irreductible a la razón o al método. En esta tradición, la obra de arte se interpreta como la manifestación de un talento innato, singular e intransferible. Romo (1997) identifica esta idea como parte del mito del genio creador, que exalta la excepcionalidad individual y excluye el componente social o aprendible de la creatividad. Esta

concepción ha alimentado una visión elitista del arte, donde el genio aparece como un elegido dotado de un talento imposible de transmitir o enseñar.

En contraste, Boden (2004, 2011) cuestiona esta mirada al proponer que el pensamiento creativo surge de procesos cognitivos, que operan en distintos niveles de complejidad combinando experiencia y conocimiento. Desde esta perspectiva, la creatividad no depende exclusivamente del don natural, sino de la interacción entre las capacidades mentales, la cultura y la práctica deliberada. El “genio” sería, entonces, una intensificación de facultades humanas compartidas, una manifestación extraordinaria de mecanismos cognitivos comunes, no una ruptura con ellos, sino una prueba de su potencial más elevado.

Ejemplos paradigmáticos de esta visión del genio creador abundan en el Romanticismo europeo. Beethoven fue considerado un espíritu casi sobrenatural, capaz de transformar el sufrimiento en belleza universal, mientras Goethe y su *Fausto* encarnaron la búsqueda incesante del conocimiento y la trascendencia individual. Lord Byron, con su vida rebelde y su obra apasionada, fue exaltado como ejemplo del artista maldito dominado por una fuerza interior incontrolable. Para Boden (2011), estas figuras consolidaron el mito del genio solitario, inspirado por una energía interior única, pero también alimentaron la idea moderna del creador como conciencia autónoma que sustituye la divinidad por el espíritu individual.

La tercera concepción mitológica sobre la creatividad artística es aquella que la entiende como una expresión de la subjetividad del artista. Romo (1997) muestra que, desde este paradigma, la creatividad es concebida como un fenómeno ambivalente que oscila entre la patología y la autorrealización. En un extremo, el acto creador se asocia a la locura, la melancolía o la alienación del artista; en el otro, se interpreta como una forma de plenitud vital y afirmación del yo. Lejos de ser opuestas, estas dimensiones convergen en el proceso creativo, donde el sufrimiento y el gozo constituyen fuerzas complementarias.

En este sentido, la creación artística puede comprenderse como una manifestación directa de la interioridad del artista, donde la obra funciona

como un medio privilegiado para exteriorizar sus tensiones afectivas, conflictos inconscientes o representaciones simbólicas de su mundo emocional (Romo, 1997). El proceso creativo se convierte así en un modo de hacer visible lo invisible, de traducir en formas sensibles aquello que habita el ámbito de la emoción y del deseo. Desde esta perspectiva, la creatividad no se explica solo por la inteligencia o la técnica, sino que se enraíza en la vida afectiva, en la capacidad del sujeto para transformar su experiencia emocional en una configuración estética. El arte, por tanto, aparece como un puente entre el sentir y el crear, entre el yo interior y su expresión simbólica.

Ejemplos paradigmáticos de esta visión son aquellos artistas que transformaron su mundo interior en materia estética. Vincent van Gogh, en cartas y lienzos como *La noche estrellada*, plasmó su conflicto emocional y su búsqueda de redención espiritual mediante el color y el movimiento. Edvard Munch, con *El grito*, convirtió la angustia existencial moderna en una imagen del desgarró psicológico. Con el surrealismo, esta exploración de la interioridad se radicaliza: André Breton, Salvador Dalí o Max Ernst conciben la creación como una vía para liberar el inconsciente a través del sueño, la automatización y la asociación libre. En la literatura, Sylvia Plath y Franz Kafka también canalizaron la alienación en formas simbólicas de autoconocimiento. En todos estos casos, el arte actúa como catarsis, revelando las zonas más profundas de la subjetividad.

Resumiendo, las tres perspectivas analizadas trazan una genealogía de los mitos de la creación artística que dominaron la cultura occidental. Desde la inspiración divina hasta el culto al genio y la interioridad emocional, estas concepciones coinciden en situar la creatividad fuera del ámbito de la razón y del conocimiento verificable. Todas ellas construyen una imagen del artista como mediador entre mundos, ya sea entre lo divino y lo humano, lo extraordinario y lo común, o lo consciente y lo inconsciente, consolidando así la idea del arte como un misterio irreductible.

Sin embargo, comprender la persistencia de estos mitos es fundamental para reconocer el modo en que configuraron nuestra visión actual del arte y del artista. A partir de este recorrido, se vuelve posible abrir

el camino hacia una lectura contemporánea de la creatividad que supere la oposición entre inspiración y racionalidad. La siguiente sección abordará precisamente este tránsito: cómo la investigación científica del siglo XX comenzó a desmontar estos imaginarios, proponiendo modelos que explican el acto creativo como un proceso cognitivo, cultural y tecnológico, condición necesaria para pensar hoy la creatividad en la era de la inteligencia artificial.

La secularización de la creatividad: del mito al análisis neuronal

El estudio científico de la creatividad ha transitado un largo camino desde las especulaciones metafísicas hacia su consolidación como objeto legítimo de investigación. En el siglo XX, la psicología fue el primer campo en otorgarle estatuto teórico y metodológico, abordándola desde diversas corrientes como son el psicoanálisis, el conductismo, humanismo y psicología cognitiva. Posteriormente, el desarrollo de las neurociencias permitió trasladar la pregunta por la creatividad al nivel biológico, buscando en la arquitectura cerebral los mecanismos que posibilitan la generación de ideas nuevas. Este recorrido, que va del estudio del comportamiento y la motivación al análisis de las redes neuronales, muestra cómo la creatividad pasó de ser una noción difusa a constituirse en un fenómeno complejo, interdisciplinario y objetivable dentro del campo científico contemporáneo.

Las perspectivas psicológicas clásicas sobre la creatividad abarcan el psicoanálisis, el conductismo y el humanismo. Desde el psicoanálisis, la creatividad se interpretó inicialmente como sublimación de impulsos inconscientes (Freud, 2013), aunque autores posteriores como Paul Matussek la vieron como expresión de la fortaleza del yo e integración de tensiones internas, considerándola una vía de autoconocimiento y transformación psíquica. Desde el enfoque conductista, la creatividad se atribuye a factores externos, enfocándose exclusivamente en comportamientos observables y dejando de lado los fenómenos mentales. Según Skinner (1969) la creatividad artística no es un proceso mental interno, sino una conducta moldeada y mantenida por las contingencias de refuerzo del ambiente social y estético.

De hecho, los defensores de esta corriente argumentan que los individuos se involucran en actividades creativas, en gran medida, por las

recompensas materiales que estas les ofrecen (Romo, 1997). Finalmente, la psicología humanista redefinió la creatividad como una capacidad universal ligada a la autorrealización, asociándola con la plenitud existencial y la salud mental, donde crear implica autenticidad, apertura y conexión con la experiencia inmediata (Romo, 1997).

No obstante, es importante señalar que, si bien estas corrientes de pensamiento abordaban tangencialmente la creatividad, no la establecían como un objeto de estudio independiente y sistemático. Por el contrario, la creatividad se integraba y diluía como un componente más dentro de un marco conceptual y metodológico mucho más amplio. Esto significaba que su análisis carecía de la profundidad necesaria para comprender plenamente sus intrincadas dinámicas, limitando así el desarrollo de teorías y modelos específicos para su estudio.

El trabajo del psicólogo Guilford es de importancia fundamental para el estudio de la creatividad, siendo uno de los primeros científicos en abordarla formalmente. El hito decisivo fue su conferencia de 1950 ante la *American Psychological Association*, en la cual reclamó que la atención científica se centrara en la creatividad. Este llamado, al señalar que los psicólogos trataban el tema de manera superficial, provocó un rebote en el campo de investigación y un rápido ascenso de las publicaciones sobre creatividad en las décadas posteriores (Romo, 1997).

Guilford (1950), como psicometrista, estableció una base conceptual y psicométrica al clasificar la creatividad como una operación del intelecto humano denominada “pensamiento divergente”. Su enfoque diferenció la creatividad de la inteligencia, y lo impulsó a desarrollar una serie de pruebas estandarizadas que buscaban medir factores como la fluidez ideacional, la originalidad y la flexibilidad. Estas pruebas han inspirado gran parte de los intentos posteriores de medir la creatividad, constituyendo la base de la tradición psicométrica del campo.

A partir de los años ochenta, la creatividad se interpreta como un fenómeno que emerge de la interacción entre individuo, cultura y contexto. Mihály Csikszentmihalyi (1996) propuso el modelo de sistemas, que define la

creatividad como el resultado de la relación entre la persona, el campo y el ámbito (Gardner, 1997; Romo, 1997). Años más tarde, Howard Gardner planteó una comprensión multinivel de la creatividad, articulando dimensiones biológicas, cognitivas y socioculturales. Según Gardner (1997), comprender la creatividad requiere atender a los niveles subpersonal (neurobiológico), personal (psicológico), impersonal (epistémico) y multipersonal (social). Esta perspectiva se refleja en su análisis de figuras históricas, donde examina cómo los contextos y los sistemas simbólicos condicionan la expresión creativa del sujeto (Gardner, 1997).

Hasta aquí hemos revisado como es el tratamiento que, desde la psicología, se le dio a la creatividad, a continuación, se expondrá las nociones principales que tienen las neurociencias.

La investigación contemporánea en neurociencias ha transformado la comprensión de la creatividad, desplazándola del ámbito del mito hacia el estudio empírico del cerebro. Lejos de ser un don misterioso, la creatividad se concibe como una función compleja que emerge de la interacción dinámica entre redes neuronales, procesos cognitivos y estados afectivos (Abraham, 2012). Esta perspectiva sostiene que la capacidad de generar ideas originales y adaptarlas a distintos contextos depende de mecanismos biológicos que integran imaginación, memoria, emoción y voluntad.

Los estudios de Abraham muestran que la creatividad no reside en una única región cerebral, sino que resulta de la cooperación entre distintos sistemas. Mediante técnicas de neuroimagen (EEG, fMRI), se ha observado la activación conjunta de áreas frontales, parietales y temporales durante tareas creativas, lo que evidencia la interacción entre el pensamiento divergente y el pensamiento convergente, encargado de evaluarlas y seleccionar las más adecuadas. Este equilibrio entre imaginación y racionalidad sustenta la flexibilidad cognitiva del pensamiento creativo.

Por su parte, Carlsson (2012) sugiere que la creatividad se asocia a una desinhibición cortical moderada, que permite conexiones inusuales entre ideas y estimula la comunicación entre hemisferios cerebrales. Las personas creativas muestran una mayor capacidad para alternar entre el razonamiento

lógico y el pensamiento analógico, lo que explica tanto su originalidad como cierta vulnerabilidad emocional. En esta línea, Colom (2012) destaca que la creatividad no se limita al arte, sino que interviene en la resolución de problemas, la innovación científica y la adaptación social. Aunque la inteligencia proporciona las estructuras de razonamiento, la creatividad posibilita romper esquemas y generar soluciones nuevas; ambos procesos comparten redes frontoparietales implicadas en la integración de la información.

La neuroestética, amplía este enfoque al campo artístico, mostrando que el cerebro del creador integra emoción y cognición de manera singular. La experiencia estética activa regiones como la corteza orbitofrontal medial y el sistema límbico, áreas vinculadas al placer y la recompensa, lo que sugiere que la belleza y la creación tienen una base neurobiológica compartida con otras formas de gratificación. La neurociencia revela que la creatividad no es un misterio insondable, sino una manifestación natural del cerebro humano, donde emoción, pensamiento y percepción convergen para dar forma a la invención artística y al conocimiento.

En conjunto, los enfoques psicológicos y neurocientíficos configuraron un giro decisivo en la comprensión moderna de la creatividad. Al desplazarla del terreno de lo mítico, lo divino o lo trascendente donde durante siglos fue concebida como don o inspiración, estas disciplinas la redefinieron como un fenómeno cognitivo, emocional y biológico susceptible de análisis empírico. La psicología permitió situarla en el ámbito del comportamiento, la motivación y los procesos mentales, mientras que las neurociencias la inscribieron en la materialidad del cerebro, en sus redes neuronales y dinámicas fisiológicas. De este modo, la creatividad dejó de ser un privilegio excepcional para convertirse en una capacidad constitutiva de la mente humana, observable y, en cierta medida, mensurable. Este desplazamiento hacia la objetividad científica no eliminó su dimensión enigmática, pero sí la reinscribió en un marco explicativo que combina lo psicológico y lo biológico.

Esta nueva racionalidad prepara el terreno para la siguiente sección, donde la teoría de Margaret Boden retomará esta herencia científica para proponer un modelo de creatividad computacional, capaz de articular los mecanismos de la mente humana con los sistemas artificiales que buscan emularla.

El giro cognitivo: un marco teórico para la creatividad artificial

Las secciones anteriores abordaron la creatividad desde dos perspectivas: primero, la del mito del arte como inspiración trascendente, como genio artista y como expresión de la subjetividad; luego, la de la psicología y la neurociencia, que buscan explicar los mecanismos cognitivos y neuronales del acto creativo. Sin embargo, ambas aproximaciones resultan insuficientes para comprender cómo la creatividad puede ser también modelada, simulada o expandida mediante sistemas artificiales.

En este punto, las ciencias cognitivas, y especialmente la obra de Margaret Boden, ofrecen un puente conceptual. Su propuesta integra los enfoques cognitivos y computacionales en una teoría que permite pensar la creatividad no como una esencia exclusivamente humana, sino como un proceso que puede representarse formalmente. A partir de su enfoque, esta sección explora cómo las ciencias cognitivas redefinen la frontera entre pensamiento humano y algoritmo, abriendo un campo de indagación donde arte, mente y máquina convergen.

Según Robles Aguirre et al. (2015), las ciencias cognitivas constituyen un campo necesariamente interdisciplinario que busca comprender cómo la mente humana percibe, representa, aprende, imagina y crea conocimiento. Surgidas a mediados del siglo XX, integran aportes de la psicología, la neurociencia, la filosofía de la mente, la lingüística, la antropología, la inteligencia artificial y la computación. Esta convergencia permite analizar los procesos mentales desde distintos niveles. Desde la psicología se estudian los mecanismos de pensamiento y aprendizaje; desde la neurociencia, las bases cerebrales que los sustentan; y desde la inteligencia artificial, se modelan estos procesos mediante sistemas computacionales. Así, las ciencias cognitivas ofrecen una visión integrada de la cognición como fenómeno emergente de

la interacción entre cerebro, mente y entorno. En el estudio de la creatividad, este enfoque permite vincular la actividad neuronal con los procesos psicológicos y las representaciones simbólicas que intervienen en la producción de ideas nuevas.

Margaret A. Boden es una destacada investigadora británica en el campo de la ciencia cognitiva y la filosofía de la mente, reconocida por sus estudios sobre la creatividad y la inteligencia artificial. Boden combina enfoques de la psicología, la informática y la filosofía para analizar los procesos mentales desde una perspectiva interdisciplinaria. Su propuesta resulta especialmente relevante para los estudios sobre creatividad computacional porque ofrece un marco teórico sólido, ampliamente aceptado por distintos autores del campo (du Sautoy, 2019; Paul & Kaufman, 2014; Zylinska, 2020). Su enfoque integra dimensiones psicológicas, culturales y cognitivas, lo que permite analizar la creatividad tanto en humanos como en sistemas de inteligencia artificial. Esta perspectiva compartida facilita una base conceptual común desde la cual evaluar si las máquinas pueden considerarse creativas, y bajo qué condiciones sus producciones alcanzan valor, novedad y sentido estético o cultural.

Para Boden (2004, 2011), crear implica producir algo nuevo, valioso y sorprendente, tres criterios que articulan una definición operativa ampliamente aceptada y que ayudan a distinguir la simple invención técnica de la auténtica creación. "Nuevo" alude a la capacidad de generar ideas o productos originales que no existían antes; "valioso" significa que esas creaciones tienen una relevancia estética, cultural o funcional reconocida por una comunidad; y "sorprendente" hace referencia a su capacidad para desafiar expectativas, provocar asombro o abrir nuevas formas de percepción y comprensión.

Boden amplía esta definición al proponer dos niveles de creatividad: la P-creatividad y la H-creatividad. La primera se refiere a la novedad psicológica o individual, es decir, a cuando una persona descubre o inventa algo nuevo para sí misma, aunque ya existiera en la historia. La segunda, en cambio, apunta a la novedad histórica o cultural, cuando la creación

representa un avance real en el conocimiento o la expresión artística de una sociedad. Esta distinción permite comprender cómo el proceso creativo opera tanto en la experiencia personal del creador como en la evolución colectiva de la cultura. En ambos niveles, la creatividad no debe entenderse como un acto aislado o espontáneo, sino como un proceso contextual y socialmente validado que depende de marcos de significado compartidos (Paul & Kaufman, 2014).

La autora identifica tres formas fundamentales de creatividad: combinacional, exploratoria y transformacional, cada una vinculada a diferentes modos de innovación cognitiva y cultural. La “creatividad combinacional” consiste en recombinar ideas o conceptos ya existentes de formas inesperadas, generando asociaciones nuevas que amplían el sentido de lo conocido. La “creatividad exploratoria” implica examinar y expandir las posibilidades de un sistema o dominio conceptual sin alterar sus reglas básicas, como ocurre cuando un artista desarrolla variaciones dentro de un mismo estilo. Finalmente, la “creatividad transformacional” representa el nivel más radical: modifica las reglas mismas del sistema y produce un cambio paradigmático que altera la forma de entender un campo completo.

Boden (2011) también aborda ejemplos de obras artísticas generadas con inteligencia artificial para ilustrar cómo los sistemas computacionales pueden expresar distintos grados de creatividad. Analiza casos de arte generativo, donde programas autónomos producen obras visuales o musicales a partir de reglas predefinidas pero con resultados imprevisibles, como los proyectos desarrollados por Harold Cohen con su sistema *AARON*. Asimismo, menciona experiencias con música algorítmica y compositores que utilizan programas capaces de aprender patrones sonoros, como David Cope con *Experiments in Musical Intelligence (EMI)*. Estos ejemplos muestran cómo la creatividad computacional no se limita a la imitación: las máquinas pueden participar en procesos de exploración y descubrimiento estético, poniendo en práctica los principios de creatividad combinacional y exploratoria definidos por la autora.

El pensamiento de Boden representa un punto de inflexión en la comprensión contemporánea de la creatividad y su relación con la inteligencia artificial. Su aporte radica no solo en haber ofrecido una definición operativa y sistemática del fenómeno creativo, sino también en haber demostrado que los procesos mentales implicados en la imaginación, la innovación o el descubrimiento pueden ser formalizados y modelados computacionalmente. Con ello, Boden introduce una perspectiva científica capaz de tender puentes entre mente, cultura y tecnología.

Su enfoque cognitivo permite reconocer que la creatividad no es un privilegio humano inefable, sino una capacidad emergente que puede estudiarse, simularse e incluso amplificarse mediante algoritmos. Esta idea inaugura una nueva forma de pensar la autoría, la producción y el valor en el arte generado con IA. Para poder comprender la singularidad de la creatividad computacional y su vínculo con lo humano, la próxima sección examinará cómo funcionan los mecanismos, estructuras y condiciones técnicas mediante los cuales los sistemas artificiales generan resultados que pueden ser considerados, al menos parcialmente, creativos.

Arquitectura de la creatividad computacional: redes neuronales, modelos generativos y aprendizaje profundo

El concepto de inteligencia artificial hunde sus raíces en la tradición de la lógica formal aristotélica, que aspiraba a representar el razonamiento mediante reglas y deducciones (Zylinska, 2020). Esa aspiración se concretó siglos más tarde con el desarrollo de la computación y los primeros programas capaces de resolver problemas siguiendo pasos lógicos explícitos. Sin embargo, los sistemas iniciales basados en reglas fijas y razonamientos deterministas demostraron rápidamente sus limitaciones para abordar tareas complejas o abiertas, aquellas que exigen adaptación, intuición o improvisación. El modelo clásico de IA, de carácter *top-down*, encarnaba una visión rígida del pensamiento, más cercana a la reproducción de procesos analíticos que a la creación genuina (Zylinska, 2020).

El giro contemporáneo hacia el aprendizaje automático (*machine learning*) y el aprendizaje profundo (*deep learning*) transformó radicalmente este

panorama. Estos modelos funcionan de manera *bottom-up*, permitiendo que el conocimiento emerja de la interacción con los datos, en lugar de imponerse mediante reglas predefinidas (du Sautoy, 2019). En términos de Boden (2004), este tránsito puede leerse como un desplazamiento desde la creatividad combinacional (la simple recombinación de elementos existentes) hacia formas de creatividad exploratoria, donde el sistema navega espacios de posibilidades dentro de un marco conceptual delimitado.

Las redes neuronales profundas son el núcleo de esta nueva etapa. Inspiradas en la estructura del cerebro humano, se componen de capas sucesivas de procesamiento que extraen progresivamente patrones y relaciones de alto nivel (Zylinska, 2020). Estas redes aprenden de grandes volúmenes de datos y ajustan de manera autónoma sus conexiones internas mediante procesos iterativos de optimización. De este modo, generan representaciones abstractas que no han sido explícitamente programadas. Esta capacidad de autoajuste, o aprendizaje no supervisado, aproxima a las máquinas a lo que Boden (2011) denominaría creatividad exploratoria: un sistema que, dentro de reglas internas, descubre configuraciones novedosas que sorprenden incluso a su programador.

Para Astobiza (2022), un caso paradigmático de esta dinámica es el de las redes generativas antagónicas (GAN), compuestas por dos subredes que aprenden en oposición dialéctica: una red generadora produce imágenes, sonidos o textos, mientras la red discriminadora evalúa su coherencia o autenticidad. A través de la competencia y retroalimentación entre ambas, el sistema progresa hacia resultados cada vez más sofisticados y difíciles de distinguir de los creados por humanos. Este proceso puede entenderse como una forma de creatividad combinacional que, en su iteración continua, se aproxima a lo transformacional: las reglas mismas del sistema se ajustan dinámicamente a medida que las redes aprenden una de otra, generando nuevas estructuras estéticas y conceptuales.

La distinción entre P-creatividad y H-creatividad propuesta por Boden (2004, 2011) ayuda a entender el lugar del artista en este nuevo ecosistema. Cuando un programador o artista trabaja con un sistema

generativo, experimenta una P-creatividad extendida: el proceso de descubrimiento se da tanto en el humano como en el algoritmo. Pero cuando los resultados se incorporan a circuitos culturales de validación (exposiciones, concursos, mercado del arte), lo que emerge es una forma híbrida de H-creatividad, donde la novedad histórica es co-producida. Esta copresencia redefine la noción de autoría y desplaza la creatividad hacia un campo de interdependencia entre agentes biológicos y técnicos.

Desde la perspectiva cognitiva, el funcionamiento de la creatividad computacional puede verse, por tanto, como una forma de simulación de los procesos de pensamiento descritos por Boden, pero con efectos que rebasan la simulación. Los algoritmos no solo imitan la mente, sino que introducen nuevas condiciones de posibilidad para la creación. En este sentido, la creatividad artificial no sustituye a la humana, sino que la expande: multiplica sus medios, acelera sus procesos y transforma su lógica. El sistema generativo funciona como un laboratorio epistemológico que pone a prueba la teoría de Boden, demostrando que los principios de novedad, valor y sorpresa pueden adquirir concreción técnica.

El funcionamiento actual de la inteligencia artificial revela una arquitectura capaz de reproducir ciertos aspectos del pensamiento creativo humano, aunque todavía de manera parcial y controlada. Los sistemas de aprendizaje profundo y de redes generativas operan mediante procesos de reconocimiento, combinación y extrapolación de datos que permiten producir resultados novedosos dentro de los límites impuestos por la información previa y los marcos algorítmicos. Esta capacidad de procesamiento, basada en la detección de patrones y en la exploración estadística de posibilidades, constituye el núcleo de lo que podría considerarse una creatividad combinatorial y, en algunos casos, exploratoria: la IA reorganiza lo existente y ensaya nuevas configuraciones dentro de su propio sistema de reglas.

Sin embargo, a pesar de su creciente sofisticación, no ha alcanzado todavía el nivel de creatividad transformacional, aquel en el que las reglas mismas del sistema se modifican y dan origen a un nuevo paradigma de

sentido. Las máquinas crean dentro de un mundo dado, pero no generan mundos nuevos. En ese contraste reside el horizonte y la frontera de la creatividad computacional contemporánea. En la próxima sección veremos cómo el uso de estas tecnologías trastoca la función del artista desde un enfoque poshumanista.

Reconfiguración del rol del artista en el escenario poshumano

El desarrollo del arte basado en inteligencia artificial (IA) obliga a replantear de raíz la figura del artista, la naturaleza del proceso creativo y las nociones modernas de autoría. Tradicionalmente, la tecnología había sido comprendida en el arte como un instrumento de ampliación, una extensión de las capacidades expresivas humanas. Sin embargo, como advierte Astobiza (2022), con la aparición de sistemas algorítmicos capaces de generar obras de manera autónoma se ha producido un desplazamiento hacia un uso sustitutivo: la máquina ya no es solo herramienta, sino agente activo de producción artística. Este tránsito implica un giro epistémico que desestabiliza el modelo humanista entendido como manifestación de una interioridad individual y subjetiva. El acto creativo se concibe, en cambio, como un proceso distribuido entre humanos y máquinas, entre inteligencias biológicas y artificiales que interactúan en redes de co-producción simbólica.

La relación entre artista y tecnología deja así de ser instrumental para volverse relacional. Zylinska (2020) propone pensar el arte de IA no en términos de oposición (humano versus máquina) sino de colaboración y co-agencia. El proceso artístico se presenta como un ensamblaje en el que ambos agentes, humano y no humano, contribuyen a la generación de sentido. La inteligencia artificial no sustituye al artista, pero transforma su función: se convierte en un medio con autonomía artificial que participa en la toma de decisiones estéticas y en la configuración de nuevas sensibilidades.

En este contexto, el creador ya no produce directamente la obra, sino que diseña las condiciones de posibilidad del sistema generativo, un espacio de “libertad programada” donde lo imprevisible puede emerger. La creatividad deja de residir en la intención individual y se desplaza hacia el entramado de relaciones técnicas, conceptuales y afectivas que componen el

dispositivo artístico. Este cambio redefine el proceso artístico como flujo continuo más que como acto puntual de creación. Las obras generadas con IA son potencialmente infinitas: se actualizan con cada iteración del algoritmo. El artista se convierte, entonces, en programador, curador o arquitecto de sistemas estéticos.

Esta redistribución de la agencia creativa genera, inevitablemente, tensiones en torno a la autoría. Astobiza (2022) señala que las nociones modernas, centradas en la figura individual del autor y en la idea de intencionalidad, se ven cuestionadas por la aparición de sistemas que operan sin supervisión directa. En las obras generadas por IA, la intervención humana puede ser mínima o incluso invisible, lo que plantea dilemas acerca de la responsabilidad y la propiedad de los resultados. La IA, al actuar como cocreadora, desafía la lógica de la autoría centralizada y promueve una concepción colectiva y procesual del acto creativo. Frente a este escenario, López de Mántaras (2017) propone medir la creatividad computacional no en términos de sustitución, sino de colaboración: las máquinas amplían las posibilidades expresivas y democratizan el acceso a la creación artística, redefiniendo la agencia humana como mediación entre lenguajes técnicos y culturales.

El arte con IA, documentado por festivales e instituciones como *Art Electronic* (Stocker, Jandl & Hirsch, 2021), configura así un nuevo ecosistema en el que confluyen ciencia, estética y tecnología. El artista contemporáneo se convierte en investigador transdisciplinar, articulando saberes provenientes del diseño, la programación, la filosofía y la teoría crítica. La automatización no elimina el trabajo artístico, sino que lo transforma: convierte el hacer manual en una operación conceptual donde programar, entrenar o curar sistemas adquiere estatuto artístico. De este modo, la práctica artística se expande hacia un territorio híbrido que combina reflexión, experimentación técnica y crítica cultural. La creatividad se vuelve un proceso colectivo y tecnológicamente mediado que obliga a repensar las fronteras entre arte, ciencia y sociedad.

Este desplazamiento encuentra su correlato filosófico en el pensamiento poshumanista, que cuestiona la centralidad ontológica del sujeto humano y propone una comprensión relacional de la inteligencia y la creación. Como sostiene Chavarría Alfaro (2015), el poshumanismo surge como respuesta a las mutaciones epistémicas introducidas por las tecnologías digitales, la cibernética y, de manera decisiva, la inteligencia artificial. Estas transformaciones no solo problematizan la frontera entre lo natural y lo artificial, sino que modifican el modo en que entendemos la agencia, la conciencia y la producción de sentido. La figura del artista deja de ser la fuente exclusiva de la obra para convertirse en un nodo dentro de ecologías complejas de co-creación simbólica.

Peter Sloterdijk (2000) interpreta este proceso como el agotamiento del humanismo letrado y de su proyecto civilizatorio. En *Normas para el parque humano*, diagnostica la pérdida de eficacia del modelo humanista, centrado en la domesticación simbólica, y plantea nuevas “antropotécnicas”, formas de autoproducción mediadas por sistemas técnicos. Desde esta perspectiva, el artista poshumano no hereda el aura del genio romántico, sino que opera como experimentador en redes de información, gestionando flujos materiales y digitales. Lo humano no desaparece, pero se reinscribe en un continuo natural-artificial donde la subjetividad se vuelve un proceso de mediación más que una esencia. El creador es, en este marco, un operador de interfaces entre datos, cuerpos y sensibilidades.

Donna Haraway (1991) lleva esta crítica a su dimensión más radical con la figura del *cyborg*, símbolo de la subjetividad híbrida de la contemporaneidad. El *cyborg* disuelve los dualismos constitutivos de la modernidad (humano/máquina, naturaleza/cultura, masculino/femenino, natural/artificial) y propone una ontología basada en la interconexión y la permeabilidad. Aplicado al arte, este paradigma implica concebir la creación como proceso transductivo en el que las tecnologías inteligentes dejan de ser simples herramientas y se convierten en co-agentes de pensamiento y sensibilidad. El artista-cyborg no “usa” la tecnología, sino que coevoluciona con ella, generando un campo de experimentación donde la agencia se distribuye y la autoría se negocia constantemente.

Por su parte, Fernández (2022) profundiza esta lectura al cuestionar la dicotomía naturaleza/cultura desde una perspectiva biológica y antropológica. Propone entender la creatividad como fenómeno eco-tecnológico, resultado de la interacción entre organismos, entornos y artefactos. Desde esta mirada, el artista contemporáneo no solo manipula tecnologías, sino que participa de un proceso de coevolución perceptiva con los sistemas algorítmicos. Las máquinas amplifican las capacidades humanas de representación y, al mismo tiempo, reconfiguran las condiciones mismas de posibilidad del gesto creativo. La obra se vuelve un campo de retroalimentación entre materia viva y datos, entre afecto y cálculo.

Chavarría (2015) vincula esta sensibilidad poshumanista con el horizonte del antropoceno, destacando que ambos marcos cuestionan la supremacía humana y promueven una ética de la coexistencia entre especies y materialidades. En este contexto, la inteligencia artificial se erige como el agente paradigmático de ese descentramiento, al introducir un régimen de producción simbólica no basado en la intencionalidad, sino en la emergencia de patrones generativos. El artista-curador o diseñador de sistemas estéticos se convierte así en mediador entre flujos de datos y mundos sensibles, organizando ecologías cognitivas donde la creatividad se distribuye entre múltiples agentes humanos y no humanos. Su práctica ya no busca producir objetos, sino articular procesos de relación y de sentido.

Esta transformación resuena profundamente con la noción de “estética artificial” propuesta por Manovich y Arielli (2024), quienes argumentan que la creatividad en la era de la IA se desplaza radicalmente desde la fabricación artesanal hacia una lógica de selección y curaduría evolutiva. Para estos autores, la IA generativa no suprime la agencia humana, sino que la eleva a un nivel meta-cognitivo: el artista deja de centrarse en la ejecución técnica del trazo para ocuparse de la navegación de espacios latentes y del filtrado crítico de variaciones infinitas. Complementariamente, Sungjin Park (2024) teoriza el surgimiento de un 'co-creativismo' que actualiza la noción de 'aura' benjaminiana: la autenticidad ya no reside en el original físico, sino en la irrepetibilidad del diálogo situado entre la instrucción humana (prompt) y la interpretación estocástica de la máquina.

Sin embargo, esta apertura hacia la co-creación no está exenta de tensiones ontológicas. Simona Chiodo (2024) aporta una perspectiva crítica fundamental al cuestionar si estos artefactos poseen la intencionalidad necesaria para ser considerados arte per se. A través de su análisis, Chiodo sugiere que la IA actúa como un espejo epistemológico que, lejos de reemplazar al sujeto, clarifica su función esencial: ante la potencia generativa de la máquina, el rol humano de dotar de sentido, propósito y contexto final se revela como el criterio demarcatorio ineludible. Así, la intervención del artista no solo es técnica, sino que se convierte en la garantía ética y semántica que distingue el arte del mero simulacro algorítmico.

De esta manera, el pensamiento poshumanista ofrece un marco conceptual fecundo para comprender la reorganización de las prácticas artísticas en la contemporaneidad digital. Al disolver las oposiciones natural-artificial y humano-máquina, este horizonte permite pensar el arte como un proceso de co-producción ontológica. La figura del artista se define como un nodo consciente dentro de una red de agencias heterogéneas que encarna una ética de la colaboración y una estética de la interdependencia. En la era algorítmica, la creatividad deja de emanar de un sujeto soberano y pasa a surgir de la interacción constante entre cuerpos, máquinas, datos y afectos. Este desplazamiento no solo redefine el arte, sino que reconfigura nuestra comprensión de lo humano mismo.

Conclusiones

A modo de cierre, este recorrido permite afirmar que la irrupción de la inteligencia artificial en el campo del arte no clausura las viejas preguntas sobre la creación, sino que las desplaza radicalmente. Si algo revela el análisis de los paradigmas —desde la inspiración divina hasta la creatividad computacional— es que la creatividad ya no puede sostenerse como la emanación exclusiva de un “interior” soberano, sino que debe comprenderse como un proceso distribuido, relacional y tecnológicamente mediado.

Para evidenciar la magnitud de esta transformación, resulta crucial contrastar dos figuras antagónicas. El rol del artista clásico, heredero de la tradición romántica y humanista, se definía por la ejecución manual y la

centralidad de la intención subjetiva. En este modelo, el creador operaba como una fuente única de sentido, transformando la materia inerte mediante un dominio técnico virtuoso para plasmar una visión interior intransferible; la obra era, ante todo, el testimonio físico de esa excepcionalidad humana.

En contraste, el artista en el escenario poshumanista con IA abandona la ejecución directa para asumir una función de orden superior: la de diseñador de ecologías generativas. Su práctica ya no consiste en dar forma a la materia, sino en gestionar flujos de información y posibilidades estocásticas. El artista deviene un curador de algoritmos y un “arquitecto de decisiones”, cuya destreza reside en la capacidad para formular las instrucciones precisas (prompts) y, fundamentalmente, en el criterio crítico para seleccionar, filtrar y dotar de sentido cultural a la exuberancia productiva de la máquina.

Esta redefinición del rol autoral conlleva profundas implicaciones estéticas, éticas y culturales que rebasan lo meramente técnico. En el plano estético, asistimos al surgimiento de una generación de artistas cuyas obras obligan a valorar el arte no por la complejidad de su producción sino por la singularidad del proceso curatorial que selecciona variables algorítmicas.

En el plano ético, la co-agencia con sistemas autónomos disuelve la responsabilidad individual y plantea nuevos dilemas sobre la propiedad y la autoría. Si la máquina contribuye activamente a la novedad, el artista humano debe asumir una nueva responsabilidad: la de ser el garante ético que inscribe esos resultados en un horizonte de sentido humano, evitando que la automatización derive en una mera acumulación de iteraciones.

Finalmente, en el plano cultural, esta transformación consolida el giro poshumanista. Aceptar que la creatividad puede ser un fenómeno híbrido implica renunciar a la última trinchera del excepcionalismo humano. La inteligencia artificial actúa aquí como un espejo antropológico: al externalizar la capacidad de crear, no nos vuelve obsoletos, sino que nos obliga a redefinir lo humano no como una esencia fija y aislada, sino como una potencia que se expande y se reinventa en su simbiosis constitutiva con la técnica.

Referencias

- Abraham, A. (2012). The neuroscience of creativity: A promising or perilous enterprise. En A. Perote Alejandre & M. Martín-Loeches Garrido (Eds.), *Creatividad y neurociencia cognitiva* (pp. 15-30). Fundación Tomás Pascual y Pilar Gómez-Cuétara.
- Astobiza, A. M. (2022). Arte y algoritmos. *Aisthesis*, 72, 282–297.
- Bianchini, F. (2023). A new definition of “artificial” for two artificial sciences. *Foundations of Science*, 28(3), 401–417. <https://doi.org/10.1007/s10699-022-09861-1>
- Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). Routledge.
- Boden, M. A. (2011). *Creativity and art: Three roads to surprise*. Oxford University Press.
- Carlsson, I. (2012). The creative personality in the light of neuropsychology. En A. Perote Alejandre & M. Martín-Loeches Garrido (Eds.), *Creatividad y neurociencia cognitiva* (pp. 31-50). Fundación Tomás Pascual y Pilar Gómez-Cuétara.
- Chavarría Alfaro, G. (2015). El posthumanismo y los cambios en la identidad humana. *Revista Reflexiones*, 94(1), 97–107.
- Chiodo, S. (2024). What AI “art” can teach us about art. *Inquiry*. Advance online publication. <https://doi.org/10.1080/20004214.2024.2395511>
- Colom Marañón, R. (2012). Inteligencia y creatividad. En A. Perote Alejandre & M. Martín-Loeches Garrido (Eds.), *Creatividad y neurociencia cognitiva* (pp. 51-70). Fundación Tomás Pascual y Pilar Gómez-Cuétara.
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. HarperCollins.
- Dierssen, M. (2015). *Neurobiología de la creatividad y la experiencia estética*. Consejo Superior de Investigaciones Científicas (CSIC).
- du Sautoy, M. (2019). *The creativity code: Art and innovation in the age of AI*. The Belknap Press of Harvard University Press.
- Fernández, H. (2009). Transhumanismo, libertad e identidad humana. *Thémata: Revista de Filosofía*, 41, 577–590. <https://revistascientificas.us.es/index.php/themata/article/view/594>
- Fernández, N. (2022). Perspectivas filosóficas sobre la “naturaleza humana” en el poshumanismo. *El banquete de los dioses: Poshumanismo y transhumanismo*, 10(1), 53–64.
- Freud, S. (2013). *Psicoanálisis del arte*. Alianza Editorial.
- Gardner, H. (1997). *Arte, mente y cerebro: Una aproximación cognitiva a la creatividad*. Paidós.
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. <https://doi.org/10.1037/h0063487>
- Haraway, D. (1991). *Ciencia, cyborgs y mujeres: La reinvención de la naturaleza*. Cátedra.

- López de Mántaras, R. (2017). La inteligencia artificial y las artes: Hacia una creatividad computacional. En *El próximo paso: La vida exponencial* (pp. 1–15). BBVA OpenMind. <https://www.bbvaopenmind.com/>
- Manovich, L., & Arielli, E. (2024). *Artificial aesthetics: Generative AI, art and visual media* [Informe en línea]. [Manovich.net](https://manovich.net/index.php/projects/artificial-aesthetics). <https://manovich.net/index.php/projects/artificial-aesthetics>
- Park, S. (2024). The work of art in the age of generative AI: Aura, liberation, and democratization. *AI & Society*, 40, 1807–1816. <https://doi.org/10.1007/s00146-024-01946-2>
- Paul, E. S., & Kaufman, S. B. (Eds.). (2014). *The philosophy of creativity: New essays*. Oxford University Press.
- Platón. (1988). *Diálogos I: Apología, Critón, Eutifrón, Ion* (J. Calonge Ruiz, E. Lledó Íñigo, & C. García Gual, Trans.). Editorial Gredos. (Trabajo original publicado ca. 380 a.C.)
- Robles Aguirre, F. A., Arias-Trejo, N., Hernández, M., & Guevara, M. Á. (2015). *Ciencias cognitivas: Una aproximación interdisciplinar*. Universidad de Guadalajara.
- Romo, M. (1997). *Psicología de la creatividad: Perspectivas contemporáneas*. Paidós.
- Skinner, B. F. (1969). Creating creative artists [Conferencia]. En *On the Future of Art*. Simposio celebrado en el Solomon R. Guggenheim Museum, Nueva York, NY, Estados Unidos. Guggenheim Museum Archives, Reel-to-Reel collection.
- Sloterdijk, P. (2000). *Normas para el parque humano: Una respuesta a la “Carta sobre el humanismo” de Heidegger*. Siruela.
- Stocker, G., Jandl, M., & Hirsch, A. J. (2021). *The practice of art and AI*. Ars Electronica Linz GmbH & Co KG.
- Zweig, S. (2023). *El misterio de la creación artística*. Sequitur.
- Zylinska, J. (2020). *AI art: Machine visions and warped dreams*. Open Humanities Press.

ARTICLES

Más allá del mercado y del Estado mínimo: La riqueza de la filosofía política liberal

Beyond the market and the minimal state: The richness of liberal political philosophy

MG. Guillermo Gastón Colella

gcolella@unq.edu.ar

Centro de Desarrollo Territorial de la UNQ

ORCID ID 0009-0005-4846-6154

Resumen: El artículo analiza el discurso de la derecha iberoamericana y su apropiación de una versión reduccionista del Liberalismo. Su objetivo general es demostrar cómo esta corriente política simplifica el Liberalismo al restringirlo a la defensa del libre mercado y la reducción del Estado, omitiendo sus dimensiones filosóficas y políticas más amplias. Las premisas argumentativas se basan en una revisión del pensamiento liberal clásico y contemporáneo. Se contrasta la visión de Juan Bautista Alberdi, quien enfatizó la limitación del poder estatal, con otros enfoques como el de Benjamín Constant, que distinguió entre libertad política e individual; Jürgen Habermas, que destaca el rol de la opinión pública y la deliberación y John Rawls, quien incorporó principios como el de justicia distributiva. Además, se analiza la tradición de la Economía Política Clásica y la Escuela Neoclásica, mostrando cómo el concepto de "mano invisible" ha sido utilizado para justificar una no intervención absoluta del Estado. El artículo concluye que el Liberalismo es una corriente de pensamiento compleja que incluye la defensa de derechos individuales, la justicia social y un equilibrio entre Estado y mercado. La apropiación ideológica que hace la derecha iberoamericana de su discurso no solo distorsiona su significado original, sino que empobrece el debate político al reducirlo a una oposición binaria entre intervencionismo y *laissez-faire*.

Palabras clave: Filosofía política liberal; Liberalismo Político; espacio público-político; Estado; Derecha Iberoamericana; Liberalismo Económico.

Abstract: This article analyzes the discourse of the Ibero-American right and its appropriation of a reductionist version of liberalism. Its general objective is to demonstrate how this political current simplifies liberalism by restricting it to the defense of the free market and the reduction of the state, omitting its broader philosophical and political dimensions. The argumentative premises are based on a review of classical and contemporary liberal thought. It contrasts the vision of Juan Bautista Alberdi, who emphasized the limitation of state power, with other approaches such as that of Benjamin Constant, who distinguished between political and individual freedom; Jürgen Habermas, who emphasizes the role of public opinion and deliberation. and John Rawls, who incorporated principles such as distributive justice. It also analyzes the tradition of classical political economy and the neoclassical school, showing how the concept of the “invisible hand” has been used to justify absolute non-intervention by the state. The article concludes that liberalism is a complex current of thought that includes the defense of individual rights, social justice and a balance between State and market. The ideological appropriation of its discourse by the Latin American right wing not only distorts its original meaning, but also impoverishes the political debate by reducing it to a binary opposition between interventionism and laissez-faire.

Keywords: Liberal political philosophy; Political Liberalism; public-political space; State; Ibero-American Right; Economic Liberalism.

Introducción

El Liberalismo ha sido históricamente una corriente de pensamiento plural, abarcando una diversidad de enfoques filosóficos y políticos. Sin embargo, en el discurso de la derecha iberoamericana contemporánea su significado se ha reducido a una interpretación que enfatiza la defensa del libre mercado y la limitación del rol estatal, omitiendo sus dimensiones más amplias y complejas.

El objetivo general de este estudio es evidenciar cómo la derecha iberoamericana ha simplificado el pensamiento liberal, reduciéndolo a una postura economicista. Por lo tanto, la tesis central es que el Liberalismo no puede ser reducido a una dicotomía simplista entre intervencionismo estatal y laissez-faire, ya que esto invisibiliza la pluralidad de fundamentos filosóficos y políticos que lo sustentan.

Para ello, se parte de la hipótesis de que el discurso de la derecha iberoamericana empobreció el debate político y distorsionó el significado original del Liberalismo al omitir su tradición filosófica, que aborda un equilibrio entre Estado y mercado, la defensa de los derechos individuales, la igualdad, la construcción de consensos por medio de la deliberación y la justicia social.

Las premisas argumentativas de este análisis se sostienen en una revisión del pensamiento liberal clásico y contemporáneo, explorando autores como: Juan Bautista Alberdi, Jürgen Habermas, Benjamín Constant y John Rawls; así como las tradiciones de la Economía Política Clásica con Adam Smith y la Escuela Neoclásica o Marginalista con autores como Hermann Gossen, Karl Menger y Léon Walras.

El artículo examina el uso que la derecha iberoamericana hace de los postulados de Alberdi (1996) que subrayó la limitación del poder estatal, del concepto de "mano invisible" (Ruiz valiente, 2006) de Adam Smith y de los postulados de la Escuela Neoclásica o Marginalista para justificar una no intervención absoluta del Estado. Para ello, se omiten deliberadamente las argumentaciones de Constant (1995) que distinguió entre libertad política y libertad individual, de Habermas (2012) que enfatizó la deliberación, la opinión pública y la sociedad civil como elementos centrales para lograr un consenso democrático y las de Rawls (1995) que incorporó principios de justicia distributiva.

Esta diversidad de enfoques dentro del Liberalismo contrasta con la tendencia de la derecha iberoamericana a presentar una visión monolítica y reduccionista, que ignora la complejidad de la tradición liberal.

La relevancia de este estudio radica en su contribución a la comprensión de cómo las corrientes políticas actuales reinterpretan y resignifican tradiciones intelectuales como insumo de discursos enmarcados en estrategias discursivas para justificar programas de gobierno que tienen como objetivos la desregulación comercial, el achicamiento estatal y políticas económicas regresivas.

Por lo tanto, al analizar el uso del Liberalismo por parte de la derecha iberoamericana, se busca aportar una perspectiva crítica que permita enriquecer el debate político y académico, reconociendo la diversidad y riqueza de la tradición liberal más allá de las simplificaciones. Esta reflexión es clave no solo para el estudio de la filosofía política, sino también para el análisis de las dinámicas políticas contemporáneas en América Latina.

Como objetivos particulares: I-, se analizará el discurso liberal de la derecha iberoamericana basado en los postulados de Juan Bautista Alberdi, de la Economía Política Liberal Clásica y de la Escuela Neoclásica o Marginalista; II-, se abordará la caída del Absolutismo y la conformación del Estado de Derecho; III-, se mencionaran las principales corrientes del Liberalismo, profundizando en el Liberalismo Político de Benjamín Constant, Jürgen Habermas y John Rawls; IV-, se desarrollará la conformación del poder político en el espacio público-político; V-, se diferenciará entre sociedad civil y Estado, remarcando la importancia de este último; VI-, se describirá la teoría de pesos y contrapesos como freno al poder público ostentado por el Gobierno y el Estado; finalmente, VII-, se presentarán las conclusiones.

El discurso liberal en la derecha iberoamericana

Discurso liberal de la derecha iberoamericana

La derecha en Sudamérica (Jeifets & Jeifets, 2016) ha mantenido una fuerte presencia en la región, incluso durante el "giro a la izquierda" de principios del siglo XXI, donde influyó en decisiones gubernamentales mediante lobby, medios de comunicación, intelectuales y think tanks.

Las crisis económicas y los escándalos de corrupción en los Gobiernos progresistas facilitaron su resurgimiento, capitalizando el malestar ciudadano con un discurso centrado en la eficiencia económica, la lucha contra la inseguridad y la corrupción. Algunos Gobiernos de derecha (Böcker Zavaro, 2021) han combinado políticas neoliberales con un discurso populista conservador, promoviendo un Estado que interviene para flexibilizar el trabajo y favorecer a las élites económicas.

Además, la derecha en la región no opera de manera aislada, sino que (Goldstein, 2022) mantiene vínculos con sus contrapartes en Europa y Estados Unidos, quienes avalan discursos y acciones en defensa del libre mercado y posturas morales conservadoras, buscando conectar con votantes de clases bajas a través de identidades nacionales, étnico-raciales o religiosas.

Es un lugar común en los discursos políticos de la derecha iberoamericana resumir las corrientes y postulados de las distintas corrientes del Liberalismo filosófico-político contemporáneo al discurso liberal que se enfrentó al Absolutismo y por el que aboga al libre mercado. Algunos ejemplos son:

En Argentina, el presidente Milei (Terrile, 2024, 19 de abril) propugno una profunda revolución liberal sin intervención estatal en la economía. En una declaración posterior, el primer mandatario argentino afirmó que (La Nación, 2024, 8 de julio) una sociedad se vuelve menos justa cuando se imponen regulaciones. Además, en su discurso de apertura de sesiones ordinarias del Congreso Nacional destacó sus objetivos de (Otalá, 2025, 1 de marzo) desregulación económica y reducción del tamaño del Estado. Por su parte, el ex-presidente Macri, en una reunión del partido PRO, subrayó (Ámbito Financiero, 2024, 1 de agosto) la importancia del Liberalismo Económico y enfatizó la lucha por la idea de libertad como antítesis del populismo.

El ex-presidente de Brasil, Jair Bolsonaro, en su discurso inaugural en el Foro de Davos, ratificó (MRE, 2019) su compromiso con políticas económicas liberales, incluidas las privatizaciones, la reducción de la intervención estatal y la promoción de la libertad individual en la economía.

El ex-presidente de Chile, Piñera, en diversos discursos entre 2018 y 2022, recalcó (Gartenlaub & Arenas, 2023) la importancia de políticas orientadas al libre mercado y a la desregulación como motores del crecimiento económico. En El Salvador, el presidente Bukele (Bergengruen, 2024, 16 de septiembre) abordó la promoción de la libertad económica individual.

En Bolivia, si bien el presidente electo Rodrigo Paz Pereira (Lamecha, 2025) sostuvo discursos moderados de una “transición suave”, su contrincante, Jorge Quiroga, propuso en la campaña un “ajuste brutal” de corte tecnocrático neoliberal. En Perú, el ex presidente Pedro Pablo Kuczynski (IDEELE, 2022) ha sido uno de los defensores y “padre del Consenso de Washington” en Sudamérica.

En Paraguay, el ex-presidente Mario Abdo Benítez (Question Digital, 2024, 9 de julio) promovió el libre comercio entre los países miembros del Mercosur. En Perú, la presidenta Boluarte (Gómez Vega, 2024, 16 de noviembre) destacó la importancia del libre comercio. Asimismo, el ex-presidente del Consejo de Ministros, Pedro Cateriano, enfatizó (Gómez Vega, 2025, 17 de marzo) la relevancia de las libertades individuales y la limitación del poder estatal. Finalmente, en Uruguay, el ex-presidente Lacalle Pou, subrayó la necesidad de (Rosario3, 2023) reducir barreras comerciales y fomentar el libre mercado en América Latina.

La posición compartida sobre la no intervención estatal y el achicamiento del Estado se encuentra en sintonía con el Liberalismo que se enfrentó al Absolutismo en siglos pasados. Encontramos una síntesis de sus preceptos en la postura enunciada por Juan bautista Alberdi, el 24 de mayo de 1880, en el discurso pronunciado en el acto de graduación de la Facultad de Derecho y Ciencias Sociales, de la Universidad de Buenos Aires. Sobre los principios del Liberalismo Económico, del libre comercio y el laissez-faire, podemos hallarlos en la Economía Política Liberal Clásica y la Escuela Neoclásica o Marginalista.

Se describen sintéticamente a continuación los postulados de Alberdi y los principios del Liberalismo Económico.

Postura de Juan Bautista Alberdi

Alberdi (1996) planteó que, en las sociedades de origen grecorromano, los ciudadanos no son verdaderamente libres, sino siervos de la patria, cuya independencia se alcanza a costa de absorber las libertades individuales, controladas por el Gobierno.

El origen de las tiranías modernas en Sudamérica (Alberdi, 1996) radica tanto en su herencia grecorromana como en su pasado de carácter español. En este contexto, *El Contrato Social* de Rousseau consolidó la idea de que el ciudadano pertenece al Estado. De este modo, los Gobiernos, al personificar la patria, han asumido el papel de representantes supremos, ejerciendo un poder omnipresente sobre la sociedad.

En cambio, la opulencia y grandeza de los pueblos del norte de América y Europa (Alberdi, 1996) no provienen del poder de sus Gobiernos, sino de la libertad y el esfuerzo de sus individuos. Su desarrollo es resultado del egoísmo bien entendido de los ciudadanos. En estas sociedades, la riqueza del Estado es fruto del trabajo libre de sus habitantes, quienes, al ser dueños de su persona y bienes, se organizan para atender sus necesidades colectivas sin depender del Gobierno.

En contraste, en los pueblos de origen latino (Alberdi, 1996) los ciudadanos esperan del Estado la solución a sus problemas. Si el Gobierno no interviene, quedan sin servicios esenciales, lo que refleja una dependencia que frena el progreso y la autonomía social.

Por otro lado, el poder absoluto del Estado sobre los individuos (Alberdi, 1996) inevitablemente lleva al despotismo del Gobierno que lo representa. La única forma de evitar un Gobierno omnipotente es limitar el poder del Estado sobre los ciudadanos.

Aunque las Constituciones de Sudamérica (Alberdi, 1996) se inspiran en las de los países libres de origen sajón, en la práctica, prevalece la idea de que el derecho del individuo debe ceder ante el del Estado. Si este último

concentra todo el poder público se convierte en un mecanismo que, tarde o temprano, genera un despotismo tiránico.

Por ende, el problema (Alberdi, 1996) no radica en un gobernante o Gobierno en particular, sino en la estructura misma del Estado, que centraliza y condensa el poder de sus ciudadanos en favor de la autoridad. Mientras la maquinaria estatal siga operando un país podrá llamarse república libre en su constitución escrita, pero en su funcionamiento real seguirá siendo una extensión del Gobierno, heredero del Absolutismo del pasado.

Ergo, el núcleo de la postura que defiende Alberdi se resume en las siguientes tres premisas:

- 1- El Gobierno, al concentrar el poder público, siempre va a ejercer un poder omnipresente y avasallante sobre las libertades individuales por medio de la maquinaria estatal, lo que inevitablemente deriva en un despotismo tiránico.
- 2- El desarrollo y la grandeza de una Nación no guarda correlación con el poder de los Gobiernos, sino con la libertad y esfuerzo de los ciudadanos que encuentran motivaciones en el egoísmo de sus intereses particulares.
- 3- La resolución de los problemas colectivos de las sociedades no debe depender de la intervención estatal, sino de una organización colectiva de la sociedad civil, lo que evita que el Gobierno y el Estado frenen el progreso individual y la autonomía social.

Economía Política Liberal Clásica y la Escuela Neoclásica o Marginalista

Economía Política Liberal Clásica y Adam Smith. En la Inglaterra del siglo XVII (Ruiz Valiente, 2006) se consolidó la etapa manufacturera dentro del desarrollo capitalista, lo que provocó profundos cambios en la propiedad y en la estructura de clases rurales, así como un notable avance de las ciencias. En el plano político, este proceso desembocó

en la Revolución Inglesa de 1648, configurando el contexto histórico que favoreció el surgimiento y la posterior evolución de la Economía Política Liberal Clásica. En este marco, “todos los economistas clásicos, confiaron en el equilibrio automático, en la no posibilidad de crisis general y defendieron el Liberalismo Económico” (Ruiz valiente, 2006, p. 24).

En *Teoría de los Sentimientos Morales* Adam Smith (Ruiz Valiente, 2006) instauró los principios que denominó "motivaciones originarias de la naturaleza humana", entre ellos: el egoísmo, la conmiseración, el deseo de libertad, el sentido de la propiedad, el hábito del trabajo y la tendencia al intercambio. Estas motivaciones conducen a la sociedad hacia su "orden natural", lo que sustenta su célebre concepto de la "mano invisible", a través de la cual cada individuo, guiado por su propio interés, contribuye involuntariamente a la armonía y el Bienestar General.

Por otra parte, Smith sostenía que (Ruiz Valiente, 2006) la función principal del Estado era garantizar la libre competencia, oponiéndose a la intervención estatal en la economía y defendiendo así el Liberalismo Económico. Según su visión, las responsabilidades del Estado debían limitarse a tres ámbitos fundamentales: I- la defensa nacional, II- la administración de justicia y III- el mantenimiento del orden interno. Además, el Estado debía encargarse de aquellas instituciones que, al no generar ganancias suficientes, no resultaban atractivas para los empresarios.

Escuela Neoclásica o Marginalista. La Escuela Neoclásica o Marginalista (Ruiz Valiente, 2006) surge alrededor de 1870 y se basa en la idea de que el individuo actúa en un contexto de necesidades ilimitadas y recursos escasos, lo que lo obliga a elegir qué bienes satisfacer según su disponibilidad.

Hermann Heinrich Gossen formuló las "leyes de Gossen", que establecen el principio de utilidad decreciente y la búsqueda de la máxima satisfacción como fundamentos del comportamiento humano. Desde una perspectiva psicológica, Karl Menger lideró una corriente que abstraía al individuo de su contexto histórico y de las relaciones de producción, presentando así las categorías económicas capitalistas como universales y atemporales. Por otro lado, la rama matemática, encabezada por Léon Walras,

se caracterizó por el uso de modelos matemáticos y postuló la libre competencia como un mecanismo que garantiza el equilibrio económico y elimina la posibilidad de crisis.

En síntesis, el Liberalismo Económico sostiene que el “equilibrio automático” es capaz de regular las crisis económicas, sociales y políticas mediante la acción de la “mano invisible”, que restablece el orden natural. Por ello, esta corriente desalienta la intervención del Estado en los asuntos económicos. Además, considera que el individuo está sujeto a necesidades ineludibles cuya satisfacción depende de la escasez de recursos, lo que obliga al individuo y a la sociedad a elegir que necesidades merecen ser satisfechas y cuales no; e, indirectamente, quienes podrán satisfacerlas más y quienes menos o en lo absoluto; lo cual queda determinado no por principios éticos y morales, sino por modelos matemáticos que garantizan el equilibrio económico, evitando crisis sociales y políticas.

Contenido del discurso monista de la derecha iberoamericana

Por lo expuesto anteriormente, se observa que el discurso monista de la derecha iberoamericana, el cual resume todas las corrientes liberales a una síntesis de la postura filosófico-política sostenida por Alberdi y a los postulados del Liberalismo Económico, plantea la desconfianza en el Estado que concentra todo el poder público, lo que ineludiblemente deviene en un despotismo tiránico estatal.

La sociedad alcanza su máximo desarrollo cuando los individuos actúan libremente en busca de sus propios intereses y el Estado no se entromete en su esfera privada. Por lo tanto, la resolución de los problemas sociales, políticos y económicos no requieren de la intervención del Estado, incluso, es algo negativo que empeora las crisis; además de ser es un obstáculo para el desarrollo de la Nación.

La resolución de dichas problemáticas depende de una organización espontánea de la sociedad civil o de mecanismos automáticos del mercado, es decir, de la “mano invisible” que llevan a un “equilibrio natural”, resultado

de la interacción de los intereses egoístas de los individuos, en un sistema de libre competencia que tiende al equilibrio.

El individuo es un ser atrapado en la esfera de sus necesidades, cuya satisfacción depende de la capacidad de intercambio y acceso a recursos. Sin embargo, dado que estos son limitados, las elecciones sobre qué necesidades pueden ser cubiertas y quiénes podrán satisfacerlas quedan determinadas por modelos matemáticos que garantizan el equilibrio económico, evitando o minimizando crisis económicas con repercusiones sociales y políticas.

Caída del Absolutismo e inicio del Estado de Derecho

Para entender la desconfianza del Liberalismo de los XVII y XVIII sobre la soberanía y el poder concentrado en el Estado, debemos tener presente que en aquella época no primaba la isonomía. El poder político y la soberanía no yacían en la ciudadanía, sino que estaban concentrados en la persona del monarca absoluto, de aquí la frase atribuida a Luis XIV “L'État, c'est moi” (el Estado soy yo). El Absolutismo (Koselleck, 2007) exigía la sumisión total del individuo, quien debía ocultar su conciencia y aceptar el juicio del soberano para poder sobrevivir. La alternativa a la obediencia era la aniquilación, cuya responsabilidad recaía sobre el propio súbdito.

Hobbes estableció que (Koselleck, 2007) el príncipe estaba por encima del derecho y era su única fuente. Como legislador y juez supremo garantizaba el orden estatal, el cual solo podía mantenerse si los súbditos aceptaban su soberanía como una necesidad moral. La teoría hobbesiana implica una escisión del individuo en dos mitades: en el ámbito privado, era libre; en el público, debía someterse por completo al Leviatán. Sin embargo, la Ilustración amplió progresivamente el espacio de la convicción individual, minando esta separación y promoviendo la idea de que el hombre debía realizarse políticamente como ciudadano.

El fin del Absolutismo (Bobbio, 2005) se aceleró en momentos de crisis y revoluciones, como la Revolución Inglesa y la Revolución Francesa. La “Teoría Contractualista”, desarrollada por pensadores como Locke y Rousseau, consolidó la idea de que el poder político debía derivar del

consentimiento de los gobernados. Esta concepción individualista afirmaba que los derechos no provenían del soberano, sino que eran previos a él y que su función era garantizar su ejercicio dentro de un marco de seguridad colectiva.

La doctrina de los derechos naturales se convirtió en el fundamento de la Declaración de Derechos en Estados Unidos de América (1776), y de la Declaración de los Derechos del Hombre y del Ciudadano de Francia (1789), donde se estableció el principio de un Estado limitado cuya función primordial era la protección de las libertades fundamentales.

Con el avance del Liberalismo (Bobbio, 2005), se consolidó el modelo de Estado de Derecho, en el que el poder está sometido a leyes generales y constitucionales. Este principio, con raíces en la tradición medieval y clásica, tomó forma en el Constitucionalismo Moderno mediante mecanismos como el control parlamentario del ejecutivo, la independencia del poder judicial y la descentralización administrativa.

La disolución del Estado estamental (Bobbio, 2005) permitió el ascenso de la Democracia Representativa, en la que el individuo, y no la corporación, se convirtió en la base de la soberanía nacional. La democracia moderna emergió en paralelo con el Estado liberal, pues ambos se fundamentaron en la afirmación de los derechos naturales del individuo.

Principales corrientes del Liberalismo Político

Principales corrientes del Liberalismo

Apartándonos del discurso monista y simplificador de la derecha iberoamericana; Funes (1997) divide al Liberalismo en tres corrientes principales:

El “Liberalismo Naturalista” (iusnaturalismo) sostiene que (Funes, 1997) los derechos y libertades individuales provienen de una legalidad pre-política y pre-social, basada en un estado de naturaleza inherente a los seres humanos. En este sentido, Bobbio (2005) destaca que el fundamento

filosófico del Estado liberal se basa en la doctrina del iusnaturalismo, que plantea la existencia de leyes naturales anteriores a la formación de cualquier sociedad, accesibles mediante la razón, de las cuales derivan derechos y deberes naturales; ergo, los individuos, por naturaleza, poseen derechos fundamentales como la vida, la propiedad privada, la libertad, la seguridad, la felicidad, etc., que deben ser respetados y garantizados por el Estado.

En cambio, el “Liberalismo Historicista” defiende que (Funes, 1997) los derechos tienen un origen tradicional y se fundamentan en las costumbres e instituciones heredadas de una Nación. Ambos modelos, iusnaturalismo e historicismo, plantean una juridicidad trascendente e inmutable.

Por otro lado, el “Liberalismo Político” considera que (Funes, 1997) los derechos y libertades derivan de principios de justicia universales, pero reconoce que su legitimidad surge de una deliberación racional y plural entre los ciudadanos. En este sentido, defiende la existencia de una “razón pública” como espacio donde se construyen colectivamente los principios de justicia, instituciones y normas que rigen la convivencia social.

A los intereses de este trabajo, solamente profundizaremos en la corriente del Liberalismo Político.

El Liberalismo Político

Habiendo visto que el Liberalismo no se resume a principios utilitaristas económicos y a la postura sostenida por Alberdi; sino que existen diversas corrientes de pensamiento liberal, desarrollaremos los fundamentos, premisas y postulados del Liberalismo Político, expuestos por Benjamín Constant, Jürgen Habermas y John Rawls.

Benjamín Constant. Según Benjamín Constant (1995) la libertad para los modernos se basa en el goce pacífico de la independencia individual y privada, la cual es inspirada por el comercio y los negocios. Esto implica no estar sujetos a la voluntad arbitraria de otros, sino solo a la ley; y abarca la posibilidad de influir en el Gobierno mediante la elección de funcionarios o la presentación de peticiones. Sin embargo, en los Estados modernos, aunque

el individuo es independiente en su vida privada, su soberanía es limitada y, a menudo, solo aparente.

En contraste, para los antiguos, (Constant, 1995) la libertad consistía en el reparto del poder entre todos los ciudadanos. Mientras que en las antiguas repúblicas ciertas instituciones restringían la libertad individual sin afectar la política, en la modernidad, la independencia individual es esencial y no debe sacrificarse en favor de la libertad política. Por ello, las restricciones que antes se aceptaban hoy no son admisibles, ya que la seguridad en el disfrute de la vida privada debe ser garantizada por las instituciones. No obstante:

La libertad individual... he ahí la verdadera libertad moderna. La libertad política es su garantía. La libertad política es por consecuencia indispensable... mis observaciones no tienden en manera alguna a disminuir el precio de la libertad política... No es a la garantía a la que hay que debilitar, es el goce el que hay que extender. No es a la libertad política a la que quiero renunciar, es la libertad civil a la que reclamo con otras formas de libertad política. (1995, p. 425)

El problema de los modernos (Constant, 1995) es que, al tener menos interés por la libertad política, a menudo descuidan las garantías que esta brinda. En este contexto, el sistema representativo permite delegar en algunos individuos la defensa de los intereses de la Nación, dado que el pueblo no siempre puede ocuparse directamente de ellos. Sin embargo, para preservar la libertad es fundamental ejercer una vigilancia constante sobre los representantes y asegurarse de que rindan cuentas, reservándose la ciudadanía el derecho de revocarlos si traicionan la confianza otorgada. De aquí que “El peligro de la libertad moderna consiste en que absortos en el goce de nuestra independencia privada, y en la consecución de nuestros intereses particulares, no renunciemos demasiado fácilmente a nuestros derechos de compartir el poder político” (Constant, 1995, p. 427).

Para subsanar dicho peligro, las instituciones (Constant, 1995) deben completar la educación moral de los ciudadanos, respetando sus derechos individuales y preservando su independencia sin interferir en sus actividades

privadas. Deben fomentar la participación en los asuntos públicos, permitir el ejercicio del poder a través del voto y las decisiones, y garantizar el derecho al control y la supervisión mediante la libre expresión. Además, mediante la práctica, deben preparar a los ciudadanos para asumir estas responsabilidades.

Jürgen Habermas. Para Habermas (2012), en la concepción liberal, el proceso democrático programa al Estado en función de los intereses de la sociedad, concebida esta como un sistema de relaciones privadas estructuradas por el mercado. La política opera como un mecanismo para canalizar estos intereses frente a un Estado con funciones administrativas. Además, la arquitectura liberal incorpora la solidaridad como una tercera fuente de integración social, junto al poder estatal y los intereses privados.

En este marco, la política se entiende (Habermas, 2012) como una lucha de posiciones orientada a obtener el control del poder administrativo. El proceso de formación de la opinión y la voluntad política se rige por la competencia estratégica entre actores colectivos que buscan mantener o adquirir poder. El éxito en este proceso se mide por la cantidad de votos obtenidos, donde los ciudadanos expresan sus preferencias al elegir personas y programas políticos en función de sus intereses y expectativas.

Por otro lado, el estatus de los ciudadanos (Habermas, 2012) se define por la cantidad de derechos subjetivos que poseen frente al Estado y a otros ciudadanos. Estos derechos garantizan un ámbito de elección libre de coacciones externas, frente a restricciones estatales y a la interferencia de terceros. Los derechos políticos, bajo esta misma lógica, permiten a los ciudadanos influir en la administración estatal mediante elecciones y la formación del Gobierno, asegurando que el poder se ejerza en su beneficio como sujetos privados. El orden jurídico, según esta perspectiva, tiene como función determinar qué derechos corresponden a cada individuo.

John Rawls. Rawls (1995) plantea que, desde una “forma Igualitaria de Liberalismo”, la justicia se basa en que:

- A. Toda persona tenga derecho a un sistema equitativo de derechos y libertades fundamentales, diseñado de manera que sea válido

para todos. Dentro de este marco, las libertades políticas deben garantizarse.

- B. Las desigualdades sociales y económicas solo son aceptables si están asociadas a cargos y funciones accesibles para todos en igualdad de oportunidades, y si generan el mayor beneficio posible para los sectores más desfavorecidos de la sociedad.
- C. Las necesidades básicas de los ciudadanos deben ser satisfechas al menos en la medida necesaria para que puedan comprender y ejercer plenamente sus derechos y libertades.

Esta postura se sustenta en tres elementos clave: (a) la garantía efectiva de las libertades políticas, (b) una igualdad de oportunidades real y (c) el "principio de diferencia", que exige que cualquier desigualdad contribuya al bienestar de los más desfavorecidos.

Sin embargo, Richard Rorty (1998), desde una posición etnocéntrica y pragmática, pone en tela de juicio la idea de que todos los seres humanos tienen el mismo valor. En momentos de crisis las lealtades hacia los grupos cercanos se estrechan, mientras que los lazos con los lejanos se debilitan, lo que plantea el dilema de contraer el círculo por lealtad o expandirlo por justicia. En este sentido, para Rorty es prioritario determinar quiénes son los importantes y quiénes no.

En relación al “Liberalismo Igualitario” y a las desigualdades sociales y económicas, Rawls no es el primero en plantearlas. Colángelo (1992) explica que la igualdad es el eje central del pensamiento de Jean-Jacques Rousseau y el criterio esencial que define la vida en sociedad. Mientras que la libertad es un concepto dinámico que se ajusta a los límites de la igualdad, esta última es la condición sine qua non para garantizar la libertad de todos. Rousseau distingue entre la desigualdad natural o física y la desigualdad moral o política, que surge de privilegios sociales y económicos. El autor del *Discurso sobre el origen y fundamento de la desigualdad entre los hombres* considera ilegítimas las condiciones sociales que generan dominación y privilegios en perjuicio de otros.

Por otra parte, Rousseau (Colángelo, 1972) critica cómo la reivindicación del mérito personal puede encubrir ventajas derivadas de estructuras sociales desiguales. La desigualdad no solo crea injusticia material, sino que profundiza la dependencia mutua mientras genera enemistad. También diferencia entre necesidades reales y artificiales: las primeras son universales y urgentes, mientras que las segundas refuerzan la opresión de algunos sobre otros. La desigualdad social, al concentrar oportunidades y limitar el desarrollo equitativo del talento, perpetúa las distancias entre individuos.

Retomando a Rawls (1995), en este marco de “Justicia Distributiva”, debe primar la cooperación social guiada por reglas públicas y procedimientos que los participantes acepten. Esto implica condiciones que cada partícipe pueda aceptar siempre que los demás también lo hagan; basados en un principio de reciprocidad, permitiendo que cada uno, sean actores individuales o colectivos, persiga sus propios objetivos dentro del marco cooperativo.

Por otra parte, la “justicia como imparcialidad” (Rawls, 1995) requiere que los términos justos de cooperación sean acordados por ciudadanos libres e iguales dentro de condiciones equitativas. Para garantizar la validez de los acuerdos, se deben excluir amenazas, coerción, engaño y cualquier ventaja en la negociación. La “posición original”, basada en el “velo de la ignorancia”, busca eliminar las desigualdades derivadas de factores sociales, históricos y naturales, asegurando así un punto de partida justo.

La conformación del poder político en el espacio público-político la época contemporánea

En este punto es pertinente preguntarse ¿Cómo se construye el poder político en el espacio público-político en la democracia moderna?

Para Hannah Arendt (2009), la “acción” (diferenciada de labor y de trabajo) es la actividad entre-los-hombres, sin mediación de objetos, y refleja la pluralidad humana, esencial para la vida política, ya que todos los aspectos de la condición humana están relacionados con la política.

Es en la acción colectiva y el discurso entre-los-hombres (Arendt, 2009) que se genera el poder político y desaparece cuando se dispersan. Este no puede almacenarse ni reservarse para emergencias, como los instrumentos de la violencia, solo existe en su ejercicio real. Es lo que mantiene viva la esfera pública, el espacio donde los hombres pueden actuar y hablar juntos. La única condición material esencial para generar poder político es la unidad del pueblo. En la vida humana, la única alternativa al poder político es la fuerza, que algunos pueden ejercer a través del monopolio legítimo de la violencia. Sin embargo, aunque la violencia puede destruir el poder, nunca puede reemplazarlo.

El espacio público-político (Habermas, 2010) actúa como un receptor de problemas que el sistema político debe abordar, ya que no pueden resolverse en otro ámbito. Este espacio percibe, identifica y tematiza de manera convincente los problemas, ejerciendo presión para que sean asumidos y elaborados en el ámbito estatal.

El núcleo institucional del espacio público-político (Habermas, 2010) radica en una trama asociativa basada en la sociedad civil, que incluye asociaciones y movimientos surgidos espontáneamente. Estos recogen y amplifican los problemas que resuenan en la vida privada, transmitiéndolos al espacio de la opinión público-política. Así, la sociedad civil institucionaliza discursos y soluciones para cuestiones de interés general dentro de los espacios públicos organizados, a ser resueltos desde el Estado.

A diferencia de Habermas, Joseph Schumpeter (1996) sostiene que la voluntad política no puede basarse en una idea de bien común, ya que los individuos tienen distintas nociones de lo que esto significa. Al no sentirse directamente responsables de lo político, no invierten el esfuerzo necesario para formarse un juicio adecuado, lo que los hace vulnerables a la influencia de la mayoría o a las "ofertas" de grupos de interés con fines particulares.

Esta falta de capacidad crítica de la opinión pública (Schumpeter, 1996) permite que la voluntad política se construya artificialmente, con grupos de intereses ofreciendo sus propuestas al mercado político. La sociedad actúa como un comprador que elige los intereses que le convienen.

En este proceso, aunque el pueblo elige al Gobierno, es este último el que toma las decisiones políticas y ofrece los intereses a tratar, creando artificialmente la voluntad política a través de la competencia por el voto y el caudillismo, lo que centraliza al Presidente en la toma de decisiones.

Sociedad Civil y Estado en Hegel

Sabiendo que el poder político se construye en el entre-los-hombres por medio de la acción y el discurso; y que el espacio público-político es una caja de resonancia de las problemáticas sociales, para luego ser abordadas en el ámbito estatal, debemos diferenciar a la sociedad civil del Estado, y describir a este último.

Hegel (1988) concibe a la sociedad civil como el espacio intermedio entre la familia y el Estado, donde los individuos, movidos por sus intereses particulares, dependen unos de otros para su satisfacción. Esta interdependencia genera un “sistema de necesidades”, en el que el bienestar y los derechos de cada persona están vinculados al de los demás. Aunque los ciudadanos actúan buscando su propio beneficio, este solo puede realizarse dentro de un marco universal, es decir, mediante la mediación de la sociedad y sus estructuras.

La sociedad civil (Hegel, 1988) se configura en un ámbito basado en la necesidad y el entendimiento, donde la subsistencia y la seguridad jurídica dependen de una red de relaciones mutuas. Los individuos, al perseguir sus fines egoístas, deben adaptarse a este orden universal, integrándose en un sistema que los convierte en eslabones de una cadena colectiva, asegurando así el equilibrio entre intereses privados y el bien común.

Bobbio y Bovero (1996) consideran que Hegel critica la concepción iusnaturalista de la sociedad civil, pues esta no representa la totalidad orgánica propia de un verdadero Estado. Mientras los iusnaturalistas la conciben como una asociación voluntaria de individuos, Hegel la entiende como un momento dentro del desarrollo del “espíritu objetivo”, que comienza en la familia y culmina en el Estado.

El Estado (Hegel, 1988) es la realización suprema de la voluntad racional y de la libertad. Es una unidad sustancial que tiene un derecho superior al del individuo, cuyo deber fundamental es ser parte de él. Solo dentro del Estado, entendido como espíritu objetivo, el individuo alcanza su verdad, su objetividad y su dimensión ética. El Estado no es solo un medio para la coexistencia, sino un fin absoluto de la razón, donde la libertad se hace efectiva.

La libertad (Hegel, 1988) no debe entenderse desde la perspectiva individual, sino desde la esencia misma de la autoconciencia, que se manifiesta como una fuerza superior e independiente. En este sentido, los individuos no son más que momentos dentro de un orden universal que los trasciende, siendo el Estado la instancia en la que se concreta plenamente la vida ética.

Bobbio y Bovero (1996) sostienen que para Hegel el Estado no es un ente imaginario ni un ideal, sino una realidad concreta que resuelve de manera necesaria y permanente los conflictos derivados de la lucha por la conservación. Es el ámbito donde la razón prevalece sobre los instintos y donde se alcanza una unidad orgánica del pueblo. Así, el Estado es la única solución posible a los enfrentamientos entre intereses egoístas. El Estado es una instancia superior que guía a la humanidad hacia un orden racional y armonioso.

La teoría de pesos y contrapesos esbozada en el *Federalista*

Siendo que el Estado es una entidad concreta que actúa de manera constante y necesaria para resolver los conflictos, que representa el espacio donde la razón se impone sobre los instintos e intereses particulares, es el ámbito en el que la libertad se hace efectiva y donde el pueblo alcanza una unidad orgánica; la cuestión a considerar es ¿Cómo evitar el despotismo tiránico estatal por la concentración del poder público?

Hamilton, Madison y Jay (1943), en *El Federalista*, argumentan que la separación de poderes es esencial para la libertad política, y describen la “teoría de pesos y contrapesos” como un instrumento para evitar la tiranía de un solo individuo, de un grupo, o de la mayoría.

El objetivo principal (Hamilton, Madison & Jay, 1943) es prevenir la concentración excesiva de poder político en una sola rama del Gobierno. Al dividir el poder político y permitir que cada rama controle a las otras, se reduce el riesgo de que cualquier individuo o grupo pueda abusar de su autoridad. El núcleo de la teoría reside en la distribución del Poder Gubernamental en tres ramas distintas: el Poder Legislativo, el Poder Ejecutivo y el Poder Judicial. Cada rama tiene sus propias funciones y responsabilidades, pero ninguna es completamente independiente de las otras. En cambio, cada una está sujeta a mecanismos de control ejercidos por las otras dos.

Por otra parte, el sistema de pesos y contrapesos (Hamilton, Madison & Jay, 1943) fomenta la deliberación y el compromiso en el proceso de toma de decisiones. Esto requiere que los diferentes actores políticos trabajen juntos y hagan concesiones. Sin embargo, el sistema puede conducir a la ineficiencia y la inacción gubernamental, ya que cada rama puede bloquear las iniciativas de las otras. Por otro lado, el sistema puede ser especialmente vulnerable en tiempos de polarización política, cuando las diferentes ramas del Gobierno están controladas por partidos opuestos y se niegan a cooperar.

En sintonía con las argumentaciones de Hamilton, Madison y Jay (1943), Friedrich Hayek (1978) sostiene que el Gobierno es un mecanismo deliberado que, más allá de sus formas más simples, no puede ser guiado únicamente por mandatos ad hoc del gobernante. El Estado posee normas propias de organización para establecer su estructura, fines y funciones; además de utilizarse para orientar a los funcionarios y organismos gubernamentales en la consecución de fines concretos. En una sociedad libre, estas normas constituyen la base que asegura el mantenimiento del derecho, regulando el ejercicio del poder gubernamental.

Conclusiones

La apropiación discursiva que hace la derecha iberoamericana del Liberalismo es una versión empobrecida y políticamente funcional de una tradición filosófica mucho más rica, diversa y compleja. Al reducir el Liberalismo a una defensa irrestricta del libre mercado, la autorregulación económica y la

sospecha permanente hacia el Estado, se han privilegiado postulados de Alberdi y de la Economía Política Clásica, mientras se omiten aportes fundamentales del Liberalismo Político, la teoría democrática y la filosofía social moderna. Esta selección no solo distorsiona el legado liberal, sino que constriñe el debate público a un eje binario entre intervencionismo estatal y *laissez-faire*, que relega la densidad conceptual que caracteriza a esta corriente.

El pensamiento liberal nunca ha sido una doctrina unívoca. Constant subraya la necesidad de preservar la libertad privada mediante la libertad política ejercida en instituciones políticas sólidas, capaces de garantizar la participación ciudadana y controlar a los representantes. Habermas destaca el rol de la deliberación, la opinión pública y la sociedad civil como elementos fundamentales de un orden democrático. Rawls argumenta que la justicia liberal requiere mecanismos distributivos que aseguren la igualdad de oportunidades reales y protejan a los sectores más desfavorecidos. Hegel muestra la imposibilidad de comprender a la sociedad civil sin el Estado como instancia racional que articule los conflictos derivados de los intereses particulares. Arendt recuerda que el poder político es una construcción colectiva que surge del actuar conjunto en el espacio público.

El Liberalismo incorpora una profunda preocupación por la justicia, la institucionalidad, la participación, la igualdad política y la regulación del poder. Por ello, la reducción del Liberalismo a una mera defensa del mercado representa más una operación ideológica que una lectura fiel de su tradición. El resultado es un discurso que instrumentaliza ciertas fuentes liberales para justificar programas de desregulación, achicamiento estatal y políticas económicas regresivas, mientras silencia los aspectos normativos que históricamente buscaron equilibrar libertad, igualdad y cohesión social.

Los postulados del liberalismo político poseen consecuencias profundas para la praxis política iberoamericana. Tales implicancias desmienten el discurso reduccionista de la derecha contemporánea, no solo en relación con el rol y funcionamiento del Estado, sino también respecto del lugar que ocupa la ciudadanía en la esfera pública, más allá del mero ejercicio del sufragio. En esta línea, el recordatorio de Constant (1995) sobre la

importancia de no descuidar la libertad política como salvaguarda de las libertades individuales constituye un llamado al involucramiento y a la participación activa en los asuntos públicos.

La participación en el espacio público político y la conformación de una opinión pública implica, para Habermas (2012), la construcción deliberativa del bien común entre actores individuales y colectivos. Este proceso favorece la integración social de distintos sectores sociales y grupos de interés, evitando que las “ofertas” particulares mencionadas por Schumpeter (1996) terminen moldeando una voluntad política artificial, desconectada de las necesidades colectivas. Para Rawls (1995), dicho bien común supone una cooperación social ajena a amenazas, coerción o engaño, y asentada en un marco de equidad en derechos y libertades fundamentales que pondere el bienestar de los sectores más vulnerables.

Rousseau (Colángelo, 1972) advierte que las ventajas derivadas de estructuras sociales desiguales suelen encubrirse mediante discursos que exaltan la meritocracia, encubriendo necesidades artificiales que refuerzan relaciones de dominación. En consecuencia, la realización efectiva del bien común requiere (Hegel, 1988) del Estado, que prioriza los intereses universales por encima de los particulares o sectoriales.

Para impedir que este mismo aparato estatal derive en un despotismo tiránico fruto de la concentración del poder, tal como advertía Alberdi (1996), Hamilton, Madison y Jay (1943) proponen el sistema de “pesos y contrapesos”, materializado en el sistema presidencial de todos los países iberoamericanos. Este diseño institucional distribuye el poder gubernamental entre los poderes Legislativo, Ejecutivo y Judicial, delimita sus zonas de incumbencia, los somete a mecanismos recíprocos de control y fomenta la deliberación y el compromiso entre actores políticos que, al cooperar y negociar, impiden la tiranía de un individuo, de un grupo o incluso de la mayoría.

En definitiva, el Liberalismo es una tradición filosófico-política con múltiples vertientes, que van más allá de una defensa intransigente del mercado y la reducción del Estado. Su evolución ha incorporado principios

de justicia distributiva, equilibrio institucional y deliberación democrática. La simplificación de sus principios y postulados en el discurso de la derecha iberoamericana actual representa un empobrecimiento de su legado y un desconocimiento de sus fundamentos filosóficos e históricos.

Referencias

- Alberdi, J. B. (1996). La omnipotencia del Estado es la negación de la libertad individual. En O. Terán (Ed.), *Escritos de Juan Bautista Alberdi. El redactor de la ley* (pp. 299–319). Universidad Nacional de Quilmes.
- Ámbito Financiero. (2024, 1 de agosto). Mauricio Macri, sobre el gobierno de Javier Milei en acto del PRO: “Es un desafío ayudar a quien no está dispuesto a ser ayudado”. *Ámbito Financiero*. <https://www.ambito.com/politica/mauricio-macri-junta-su-tropa-antes-que-javier-milei-detone-al-pro-n6042360>
- Arendt, H. (2009). *La condición humana*. Paidós.
- Bergengruen, V. (2024, 16 de septiembre). How Nayib Bukele’s “iron fist” has transformed El Salvador. *TIME*. <https://time.com/7015598/nayib-bukeles-iron-fist-el-salvador/>
- Bobbio, N. (2005). *Liberalismo y democracia*. Fondo de Cultura Económica.
- Bobbio, N., & Bovero, M. (1996). *Sociedad y Estado en la filosofía política moderna: El modelo insnaturalista y el modelo hegeliano-marxiano*. Fondo de Cultura Económica.
- Böcker Zavaro, R. (2021). Desarrollo, populismo y neoliberalismo: Pensar Latinoamérica. *Revista Internacional de Organizaciones*, (25–26), 51–70. <http://www.revista-rio.org>
- Colángelo, R. (1972). Igualdad y sociedad de Rousseau a Marx. En C. Lévi-Strauss et al. (Eds.), *Presencia de Rousseau*. Nueva Visión.
- Constant, B. (1995). De la libertad de los antiguos comparada con la de los modernos. *Revista de Estudios Públicos*, (59), 421–427.
- Funes, E. (1997). *Universalismo republicano, universalismo liberal* [Material de clase]. Teoría Política Contemporánea. Universidad de Buenos Aires, Facultad de Ciencias Sociales.
- Gartenlaub, A., & Arenas, R. (2023). Discurso político en tiempos de crisis: Análisis de los discursos del presidente chileno Sebastián Piñera durante su segundo mandato. *Iberoamericana – Nordic Journal of Latin American and Caribbean Studies*, 52(1), 63–75. <https://doi.org/10.16993/iberoamericana.569>
- Goldstein, A. (2022). *La reconquista autoritaria: Cómo la derecha global amenaza la democracia en América Latina* (Vol. 97). Marea Editorial.
- Gómez Vega, R. (2024, 16 de noviembre). El cierre de la cumbre Asia-Pacífico da cierto oxígeno a la presidenta Boluarte en medio de su crisis de popularidad. *El País*. <https://elpais.com/america/2024-11-17/el-cierre-de->

- [la-cumbre-asia-pacifico-da-cierto-oxigeno-a-la-presidenta-boluarte-en-medio-de-su-crisis-de-popularidad.html](https://elpais.com/america/2025-03-17/pedro-cateriano-por-que-vargas-percibio-los-abusos-de-la-dictadura-cubana-y-garcia-marquez-no.html)
- Gómez Vega, R. (2025, 17 de marzo). Pedro Cateriano: “¿Por qué Vargas Llosa percibió los abusos de la dictadura cubana y García Márquez no?”. *El País*. <https://elpais.com/america/2025-03-17/pedro-cateriano-por-que-vargas-percibio-los-abusos-de-la-dictadura-cubana-y-garcia-marquez-no.html>
- Habermas, J. (2010). *Sobre el derecho y el Estado democrático de derecho en términos de teoría del discurso*. Trotta.
- Habermas, J. (2012). Tres modelos de democracia: Sobre el concepto de una política deliberativa. *Polis*. <https://journals.openedition.org/polis/7473>
- Hamilton, A., Madison, J., & Jay, J. (1943). *El Federalista o la nueva Constitución*. Fondo de Cultura Económica.
- Hayek, F. A. (1978). *Derecho, legislación y libertad*. Unión Editorial.
- Hegel, G. W. F. (1988). *Principios de la filosofía del derecho o derecho natural y ciencia política*. Edhasa.
- Jeifets, V., & Jeifets, L. (2016). Particularidades y perspectivas del resurgimiento del “fénix de la derecha” en América Latina. *Iberoamérica*, (3), 34–60.
- Koselleck, R. (2007). *Crítica y crisis: Un estudio sobre la patogénesis del mundo burgués*. Trotta.
- Lamecha. (2025, 20 de octubre). El ascenso de Rodrigo Paz en una Bolivia en crisis. <https://lamecha.ar/bolivia-transicion-suave-rodrigo-paz-fin-del-mas/>
- La Nación. (2024, 8 de julio). Habló Milei: “La izquierda es la ideología de los ricos y poderosos”, el discurso del Presidente en Brasil. *La Nación*. <https://www.lanacion.com.ar/politica/hablo-milei-la-izquierda-es-la-ideologia-de-los-ricos-y-poderosos-el-discurso-del-presidente-en-nid08072024>
- Ministerio das Relações Exteriores. (2019, 22 de enero). *Discurso del Presidente de la República, Jair Bolsonaro, en la sesión plenaria del Foro Económico Mundial – Davos, Suiza*. <https://www.gov.br/mre/es/centro-de-contenidos/discursos-articulos-y-entrevistas/presidente-de-la-republica-federativa-de-brasil/discursos/discurso-del-presidente-de-la-republica-jair-bolsonaro-durante-la-sesion-plenaria-del-foro-economico-mundial-davos-suiza-22-de-enero-de-2019>
- Otala, J. (2025, 1 de marzo). Presidente argentino ratifica desregulación económica y apertura comercial en inicio anual legislativo. *Reuters*. <https://www.reuters.com/latam/negocio/KBJU7XS3FNNINIXWDSG2X6MG5Q-2025-03-02>
- Question Digital*. (2024, 9 de julio). Los discursos de los presidentes en la 64.ª Cumbre del Mercosur. *Question Digital*. <https://questiondigital.com/en-video-los-discursos-de-los-presidentes-en-la-64-cumbre-del-mercosur/>
- Rawls, J. (1995). *Liberalismo político*. Fondo de Cultura Económica.

- Revista IDEELE / Análisis. (2022, 3 de agosto). El ocaso de PPK: el ideólogo (artículo/ensayo con citas a entrevistas y posiciones públicas de PPK). *Revista IDEELE*. <https://www.revistaideele.com/2022/08/03/el-ocaso-de-ppk-el-ideologo/>
- Rorty, R. (1998). *Pragmatismo y política*. Paidós.
- Rosario3. (2023, 24 de enero). El duro discurso del presidente de Uruguay en la cumbre de la Celac: “Acá hay países que no respetan la democracia”. *Rosario3*. <https://www.rosario3.com/informaciongeneral/El-duro-discurso-del-presidente-de-Uruguay-en-la-cumbre-de-la-Celac-Aca-hay-paises-que-no-respetan-la-democr-20230124-0063.html>
- Ruiz Valiente, R. (2006). *Principales doctrinas del pensamiento económico*. Ediciones de la Universidad.
- Schumpeter, J. A. (1996). *Capitalismo, socialismo y democracia*. Folio.
- Terrile, S. (2024, 19 de abril). Ante los principales empresarios, Milei defendió el modelo irlandés: suba del PBI y baja de impuestos. *TN*. <https://tn.com.ar/economia/2024/04/19/ante-los-principales-empresarios-milei-defendio-el-modelo-irlandes-suba-del-pbi-y-baja-de-impuestos/>

Poderes deónticos y epistemología social: Una aproximación al testimonio a partir de la filosofía de la sociedad de John Searle

Deontic powers and social epistemology: An approach to testimony based on John Searle's philosophy of society

Felipe Álvarez
felavarez@ug.uchile.cl
Universidad de Chile

ORCID ID 0000-0002-7153-2851

Resumen: Este artículo explora la posibilidad de un cruce sistemático entre filosofía de la sociedad de John Searle y la epistemología social a partir del testimonio y los poderes deónticos asociado al intercambio testimonial. Para dar cuenta de ello, se considera al testimonio un fenómeno epistémico cuya práctica se corresponde el deber de constituirse el agente epistémico como un garante de la verdad al testificar, de testificar honestamente y de ser capaz de reevaluar el estatuto epistémico del propio testimonio al compartirlo con otro. Para ello, se procede de la siguiente manera: en primer lugar, se da cuenta de los elementos que componen la ontología social searleana para dar cuenta de su filosofía de la sociedad. Se explicita la relación entre la asignación de las funciones de estatus y los hechos institucionales, y la instauración de reglas constitutivas, generando con ello el surgimiento de los poderes deónticos asociados a nuestros roles sociales; en segundo lugar, analizan las definiciones tradicionales del testimonio en la epistemología social, dando cuenta de la importancia de la división de roles para una comprensión integral del intercambio testimonial; finalmente, se proponen tres poderes deónticos en torno a la práctica del testimonio.

Palabras clave: Testimonio; poderes deónticos; epistemología social; ontología social; filosofía de la sociedad.

Abstract: This article explores the possibility of a systematic link between John Searle's philosophy of society and social epistemology based on testimony, as well as the deontic powers associated with testimonial exchange. Testimony is considered an epistemic phenomenon, and the practice of testimony corresponds to the epistemic agent's duty to act as a guarantor of truth, testify honestly, and be capable of reevaluating the epistemic status of one's testimony. To do this, the elements that comprise Searlean social ontology are examined to understand his philosophy of society the first place. The relationship between the assignment of status functions and institutional facts, and the establishment of constitutive rules is made explicit. This generates the emergence of deontic powers associated with our social roles. Second, traditional definitions of testimony in social epistemology are analyzed, considering the importance of role division for a comprehensive understanding of testimonial exchange. Finally, three deontic powers are proposed in relation to the practice of testimony.

Keywords: Testimony; deontic powers; social epistemology; social ontology; philosophy of society.

Introducción

John Searle sostiene en su filosofía de la sociedad (1995; 1998; 2012) que los hechos institucionales son creados y mantenidos a través de la aceptación, por parte de la intencionalidad colectiva, de funciones de estatus, las cuales imponen, por medio del acto de habla declarativos, la regla constitutiva “X cuenta como Y en C”. Esta caracterización de los hechos institucionales implica, claramente, un paso de la intencionalidad individual a la intencionalidad colectiva, paso en el que la asignación de tales funciones para la creación de hechos institucionales suscita el surgimiento de poderes deónticos respecto de los mismos, es decir, obligaciones que regulan el

comportamiento de la gente en torno a la realidad institucional. Ahora bien, a pesar de que el planteamiento de Searle describa cómo operamos en la realidad institucional, instituciones como el testimonio no son tratadas a lo largo de sus textos, dejando el trabajo de describir la ontología social relativa al testimonio a los epistemólogos sociales. Estos, en su necesidad de describir la realidad de su fenómeno para luego someterlo a un análisis epistemológico, generaron diversas clases de definiciones (Coady, 1992; Graham, 1997; Lackey, 2010). Sin embargo, la carencia de la precisión descriptiva del trabajo de Searle suscita que tales descripciones no den cuenta de las exigencias normativas asociadas a la institución del testimonio, exigencias las cuales deben ser ejecutadas de manera adecuada para su correcto funcionamiento y mantenimiento en realidad institucional.

Atendiendo a este problema, me propongo investigar en qué consisten los poderes deónticos asociados al testimonio como institución social, en específico, me propongo investigar cómo es que somos capaces de confiar en otras personas por medio del intercambio testimonial, atendiendo a las obligaciones de la gente al formar parte de la intencionalidad colectiva que sostiene al testimonio como institución social. En este sentido, se pretende ampliar la comprensión de los poderes deónticos como marco explicativo para precisar la comprensión del testimonio desde una perspectiva epistemológica.

Para lograr esto, pretendo establecer un vínculo sistemático entre la filosofía de la sociedad de Searle y el enfoque interpersonal en la epistemología del testimonio (Moran, 2006), pretendiendo demostrar que, al testificar, S no sólo produce un fenómeno de transmisión de conocimiento, sino que también cumple, junto con el oyente del acto testimonial, con los deberes asociados a la institución social del testimonio.

Siguiendo lo anterior, abordaré el problema de la siguiente manera: de acuerdo con el enfoque interpersonal del testimonio, el hablante adquiere al testificar un compromiso con el contenido proposicional de su aseveración a la par que el oyente, sirviendo, él mismo, como garantía de ello. Del mismo modo, su contraparte adquiere por dicha garantía el deber de exigir

explicaciones en caso de que el contenido aseverado no sea verdadero. Considero que estos deberes asociados se enmarcan dentro de la propuesta searleana en la medida de que el testimonio como institución sólo puede mantenerse si la intencionalidad colectiva concuerda y se respetan los deberes asociados a este: si el testimonio suscita un deber con la verdad y con la garantía de esta, entonces el mantenimiento de la credibilidad asociada a esta institución sólo será posible en la medida de que estos deberes sean cumplidos; de lo contrario, es destruida. De esta manera, propongo finalmente que la epistemología del testimonio puede ser incrustada en la filosofía de la sociedad de Searle.

Filosofía de la sociedad: su ontología y el lugar de los poderes deónticos

De acuerdo con la propuesta ontológica de Searle esbozada en *The construction of reality* (1995) y profundizada en obras posteriores (1998; 2010), vivimos en *un sólo mundo de partículas*. En este mundo convergen, fundamentalmente, dos clases de hechos, a saber, los hechos brutos y los hechos institucionales; por una parte, los hechos brutos corresponden a los fenómenos cuya realidad no depende del observador, como por ejemplo, la existencia de fenómenos naturales tales como ventiscas o terremotos; mientras que, por otra parte, los hechos institucionales corresponden a fenómenos cuya realidad es dependiente del observador, como por ejemplo, la existencia de fenómenos sociales tales como el valor del dinero. Supuesto esto, me concentraré en los últimos para responder la siguiente pregunta: ¿qué elementos constituyen las condiciones necesarias y suficientes para la generación de un hecho institucional y, por consiguiente, para la formación de la realidad institucional como un entramado de hechos institucionales? Según Searle, en principio, “[e]sto requiere de tres elementos. El asignamiento de funciones, la intencionalidad colectiva y las reglas constitutivas” (1995, p. 13)¹. Ahora bien, para responder nuestra pregunta debemos delimitar en qué consiste cada uno de ellos en la propuesta de Searle y cuál es su función dentro de su ontología. Por consiguiente, procederé a continuación a dar cuenta de ello atendiendo

¹ Para trabajar he utilizado los textos en su idioma original. Las traducciones me pertenecen.

al modo en que Searle caracteriza estos elementos en relación con la construcción de la realidad institucional. Comenzaré, por lo tanto, con la *asignación de funciones*.

Asignación de funciones de estatus

Searle distingue entre funciones no-agentivas y funciones agentivas: las primeras corresponden a las funciones que *descubrimos* en los objetos. Tal y como señala Searle, “[a]lgunas funciones no son impuestas en los objetos para servir a propósitos prácticos, sino que son asignadas a los objetos naturales y a [sus] procesos como parte de un razonamiento teórico del fenómeno en cuestión” (1995, p. 20). Ejemplo de esto puede ser el funcionamiento de un órgano (tal como lo es el funcionamiento del corazón en el ejemplo del mismo Searle) o el funcionamiento de la lluvia en el ciclo hidrológico. En ambos casos, simplemente describimos *cómo es que funciona* un fenómeno en relación con otro, sin ser nosotros mismos aquellos que estipulamos la función del objeto; por el contrario, las funciones agentivas son “[...] funciones que nosotros no descubrimos, y que no ocurren naturalmente, sino que son relaciones asignadas a los intereses prácticos de agentes conscientes” (1995, p. 20). Ejemplo de esto puede ser identificar y estipular la función del objeto ‘silla’ en un objeto determinado que no constituye una silla como tal, como bien puede serlo un tronco o una roca (siguiendo, *mutatis mutandis*, el ejemplo del mismo Searle), pues *puede funcionar* como una silla en un contexto determinado. De este modo, dotamos al objeto determinado una función que no descubrimos en él, sino que nosotros mismos somos quienes imponemos la función al objeto.

De las funciones agentivas se desprende una subclase: las denominadas funciones de estatus. Ejemplos de esta subclase de funciones son el dinero, la documentación civil e, inclusive, la asignación de función ‘entrada de cine’ al objeto ‘hoja de papel impresa’. Estas funciones son fundamentales para la creación de hechos institucionales y se distinguen de las meras funciones agentivas en la medida de que estas últimas no requieren de la aceptación de un colectivo en su instauración y mantención mientras que las funciones de estatus sí lo requieren. Los detalles a propósito de la

diferencia entre las funciones agentivas y las funciones de estatus quedarán claros una vez que tratemos distinción entre *I-intentionality* y *We-intentionality* en el siguiente apartado. No obstante, podemos precisar de momento sobre estas funciones que, tal y como señala Searle al respecto: “[e]l desempeño de una función requiere que haya un estatus reconocido colectivamente tal que la persona u objeto tiene, y sólo en virtud de ese estatus es que la persona u objeto puede desempeñar la función en cuestión” (2010, p. 7). Asimismo, presenta, de un modo similar, texto la siguiente definición:

Definiré una función de estatus como una función que es desempeñada por un(os) objeto(s), persona(s), u otra clase de entidad(es) y el cual puede ser realizado sólo en virtud del hecho de que una comunidad en la cual la función es desempeñada asigna un cierto estatus al objeto, persona o entidad en cuestión, y la función es desempeñada en virtud de la aceptación o reconocimiento colectivo del objeto, persona o entidad como poseedor de aquel estatus. (2010, p. 94)

De acuerdo con estas formulaciones, las funciones de estatus generan, a través del reconocimiento de un colectivo social, que un objeto posea una función estipulativa tal que esta sólo sea realizable en la medida de que dicho reconocimiento se mantenga vigente. Si consideramos el ejemplo anterior respecto de la función agentiva, a saber, asignar la función ‘silla’ al objeto ‘tronco’ o al objeto ‘roca’, podremos divisar que este no precisa de la vigencia del reconocimiento social para el mantenimiento de la función: en efecto, podría ser yo el único que asigne la función señalada a tales objetos sin que eso suscitara un problema. No obstante, las funciones de estatus requieren que yo, en conjunto con una comunidad determinada, designemos colectivamente la función en cuestión, pues un objeto sólo adquirirá la función estipulada si esta es reconocida y mantenida por la comunidad. De este modo, si consideramos el ejemplo anterior acerca de la entrada de cine, podremos vislumbrar que está sólo contará como ‘entrada’ en la medida en que un colectivo social sea capaz de reconocer la función asignada al objeto ‘hoja de papel impresa’, a saber, el personal del cine y los consumidores. Por lo tanto, de nada valdría que yo, en aislamiento, pretendiese efectuar tal

asignación de función si el personal del cine no opera por un sistema que permita imprimir una entrada digital para adquirir el acceso a la sala; inclusive, si lo hiciese, no me dejarían entrar por la invalidez de la asignación de función en tanto que carece del reconocimiento requerido por el personal calificado. Por consiguiente, las funciones de estatus no pueden sino mantener su vigencia sólo si estas son declaradas, aceptadas y reconocidas por un colectivo social.

Intencionalidad colectiva

Habiendo dicho esto sobre la asignación de funciones, ahora debemos distinguir entre la intencionalidad y la intencionalidad colectiva. Para empezar, debemos dejar en claro que para Searle los estados intencionales poseen la siguiente estructura:

Cada estado intencional se divide en dos componentes: el tipo de estado que es y su contenido. [...] Por ejemplo, puedo creer que está lloviendo, tener miedo de que llueva o desear que llueva. En cada caso tengo el mismo contenido proposicional, *p*, que es que está lloviendo, pero lo tengo en diferentes tipos de intención, a saber, diferentes modalidades psicológicas: creencia, miedo, deseo, entre otras, representadas por ‘*S*’. (2010, p. 27)

Asimismo, Searle señala que, de un modo similar a los actos de habla, “[e]l estado intencional representa sus condiciones de satisfacción” (2010, p. 29) como también que estos responden a un plexo de estados intencionales anteriores. Tal y como señala:

Mis estados intencionales no vienen a mí como unidades aisladas. No puedo pretender ir al cine a menos que tenga un conjunto de otras creencias y deseos. Tengo la creencia de que ir al cine son formas de entretenimiento público, que cada vez que uno ve una película lo hace por medio de ir a un cine en el cual la película está siendo expuesta [...] y así con un conjunto total de otros estados intencionales. (2010, p. 37)

Considerando estos aspectos, podemos señalar que la intencionalidad, al menos en su forma individual, consiste para Searle en encontrarse en posesión de un estado mental intencional el cual se corresponde con tipo de intencionalidad determinada y que, en la mayoría de los casos, posee contenido proposicional. Asimismo, que este responde a condiciones de satisfacción respecto de su tipo y que no surge al margen de otros estados mentales correlacionados. De este modo, por ejemplo, si poseo el estado mental ‘creo que hace calor’, este se corresponde con el tipo de la creencia y su contenido proposicional consiste en ‘hace calor’. Dado que las creencias versan sobre valores de verdad, la condición de satisfacción de mi creencia cuyo contenido proposicional es ‘hace calor’ consistirá en que mi creencia se adecue a los hechos, de modo que será una creencia exitosa sólo si esta acierta con ello. Del mismo modo, esta surgirá junto con un plexo de diversos estados intencionales relativos, tales como podrían serlo ‘debo tomar agua’ o ‘detesto sudar’, entre otros.

Habiendo caracterizado de este modo la intencionalidad, ¿cuál es la diferencia entre intencionalidad e intencionalidad colectiva? ¿es reducible la intencionalidad colectiva a intencionalidades individuales que poseen el mismo estado mental intencional paralelamente? De acuerdo con Searle, la intencionalidad colectiva no es reducible a la intencionalidad individual. Esto lo demuestra a través de sus ejemplos acerca de los estudiantes egresados de la Escuela de Negocios de Harvard, el cual consiste, en reducidas cuentas, en lo siguiente: por una parte, unos egresados creen en la teoría de la mano invisible de Adam Smith. Al día siguiente de la graduación, cada uno decide *por su propia cuenta* aplicar esta teoría en su quehacer cotidiano. Cada uno sabe que el resto está haciendo lo mismo y poseen mutuo conocimiento de sus metas. En esto no hay cooperación e inclusive hay una ideología de no cooperación operando; por otra parte, otro grupo de egresados decide el día de su graduación, *en conjunto*, seguir esta teoría y aplicarla a su vida cotidiana.

Según la propuesta searleana, en el primer caso, no hay intencionalidad colectiva, sino mera intencionalidad; cada uno de ellos decide por sí mismo cómo actuar y, aunque coincidan circunstancialmente en su decisión, esta decisión fue tomada por cada uno aisladamente. Contrario al

primero, en el segundo caso si se presenta intencionalidad colectiva debido a que la toma de decisión fue realizada por la comunidad. Tal y como señala Searle: “[e]n el primer caso, los individuos no tienen un pacto o promesa de actuar de ese modo [...] pero en el segundo caso, hay una promesa solemne hecha por cada uno para con los otros” (2010, p. 48). Esto es de suma importancia porque a la intencionalidad colectiva se le suman tres características fundamentales que no se encuentran en intencionalidad individual, a saber, la *cooperación*, la necesidad de lenguaje y el *trasfondo*: por una parte, la cooperación implica conocimiento o creencia común en la realización de actividades en conjunto; el lenguaje, la capacidad de realizar actos de habla en relación con ese conocimiento o creencias común; finalmente, el trasfondo (es decir, el conjunto de estados intencionales no conscientes que poseemos), posibilita comprender que, por ejemplo, las promesas una vez hechas no han de romperse por las condiciones sociales que estas implican u otros datos culturales relevantes al operar en la realidad institucional.

En los casos analizados, encontramos sólo en el segundo grupo estas tres características expresadas. Por ejemplo, encontramos cooperación en la medida de que hay una creencia común, a saber, ‘actuar de modo x’ que es asumida por todos los miembros de la comunidad y que sirve para la realización de esa actividad en conjunto; encontramos lenguaje en la medida de que existe un acuerdo enunciado para llevar a cabo la actividad del modo que ellos dispusieron; y encontramos de trasfondo en la medida de que comprenden que, por ejemplo, el acuerdo no debe ser roto por ninguno de ellos, sin que esto sea enunciado explícitamente. Por consiguiente, al poseer todas estas características, el grupo dos opera de acuerdo con una intencionalidad colectiva mientras que el grupo uno opera de acuerdo con un conjunto de intencionalidades individuales.

Reglas constitutivas

Ya señalado el modo en que Searle caracteriza la intencionalidad colectiva, debemos ahora señalar en qué consiste las reglas constitutivas de los hechos institucionales. Como hemos visto anteriormente, las funciones de estatus

requieren de la aceptación de un colectivo social para su creación y mantención. Sin embargo, ¿qué implica que las funciones de estatus deban ser aceptadas por dicho colectivo? ¿qué sucede realmente cuando se produce tal aceptación? De acuerdo con el planteamiento de Searle, lo que se produce es la imposición de *reglas constitutivas*.

Antes de entrar en qué consiste tal regla, debemos diferenciarla de otra clase de reglas que operan en la realidad institucional, a saber, las reglas regulativas. Según Searle, “[a]lgunas reglas regulan actividades ya existentes [...] Sin embargo, algunas reglas no meramente regulan, sino también crean la mera posibilidad de algunas actividades” (1995, p. 27). El ejemplo predilecto de Searle respecto de las primeras son las reglas del tránsito. Estas reglas regulan la actividad de conducir, siendo esta actividad anterior a las reglas mismas y no-dependiente de las mismas reglas para existir. Esto queda en manifiesto si consideramos que, de país en país, las reglas del tránsito pueden presentar ligeras variaciones las cuales no irrumpen la actividad de conducir convirtiéndola en otra actividad, sino que esta se mantiene tal cual siendo que la posibilidad de acción se ve reducida de alguna u otra manera (generalmente en el sentido jurídico, como la conocida, precisamente por su carácter polémico, prohibición ‘las mujeres no pueden conducir’ en rige en Arabia Saudita o la ley francesa que dicta que es requerido llevar un alcoholímetro desechable al conducir), más no por esto es necesario para la acción misma el atender a la regla.

Por el contrario, el ejemplo predilecto de respecto de la segunda es el juego del ajedrez. En el ajedrez sólo hay reglas constitutivas; en efecto, ninguna regla regula alguna acción ya existente, sino que delimitan entre sí qué es el juego mismo y las posibilidades de acción que el juego admite. Por ejemplo, la regla que delimita que en el ajedrez la pieza ‘caballo’ es la única pieza que puede moverse en forma de L a lo largo de cuatro escaques y que puede saltar piezas intermedias en el tracto de su movimiento, es constitutiva del juego y hace posible el juego mismo. Esto debido a que, si yo moviese el caballo como otra pieza o, si se me ocurriese, inventase un movimiento en particular para este, no estaría jugando ajedrez como tal, sino ‘otro juego’ (tal cual como si asignase una función agentiva al objeto). Sin embargo, en el

juego del ajedrez, el jugador posee respecto del caballo una posibilidad de acción determinada por las reglas del juego, de modo que cualquier infracción a la regla anula el juego mismo.

Otra característica importante de esta distinción es la forma que poseen ambas reglas. De acuerdo con Searle, “[...] las reglas regulativas tiene la forma ‘Haz X’, mientras que las reglas constitutivas tienen la forma ‘X cuenta como Y en C’” (2010, p. 10)². Asimismo, señala al respecto lo siguiente:

[...] [l]a regla *regulativa* [...] es fijamente Directiva. Su función es provocar cierta forma de comportamiento, y es satisfecha si ese comportamiento concuerda con el contenido de la regla. [...] Por el contrario, la regla *constitutiva* es fijamente Declarativa. [...] Hace de algo el caso representado ese algo como siendo el caso. (2010, p. 97)

Si volvemos a los ejemplos anteriores, podemos notar que, en primer lugar, la regla regulativa del tránsito que dicta que se debe llevar un alcoholímetro desechable al conducir en Francia constituye un enunciado cuya forma corresponde al imperativo ‘lleva contigo un alcoholímetro desechable al conducir’. Sin embargo, en el segundo caso, la regla constitutiva del movimiento del ajedrez posee la forma de un enunciado declarativo, a saber, ‘la pieza ‘caballo’ *cuenta como* la única pieza que puede moverse en forma de L a lo largo de cuatro escaques y que puede saltar piezas intermedias en el tracto de su movimiento en el ajedrez’. En este último ejemplo, la estipulación de la función de la pieza caballo por la designada es relativa al contexto, a saber, el contexto del juego ajedrez. De acuerdo con la declaración de esta regla, en ese contexto determinado, el caballo constituirá lo que dicta la variable Y, en este caso, ser ‘la única pieza que puede moverse [...]’ por lo que necesariamente será de esa manera. La diferencia radicará, entonces, en que

² Quisiera señalar que, un punto importante respecto de las funciones de estatus es que la vinculación con forma de la regla constitutiva no es realmente necesaria. Esto lo deja en claro Searle al aceptar la acotación de Barry Smith respecto de los ‘freestandards Y terms’. Cfr. Smith, Barry, 2003, p. 1-33.

el primer caso la regla es impuesta a una acción que ya existe por lo que puede ser eludida eventualmente. No obstante, en el segundo caso, la regla declara qué *representa* el objeto en un contexto determinado, de modo que esta relación hace posible que un objeto represente, de acuerdo con la intencionalidad colectiva, una función asignada en particular en un contexto determinado.

Podere deónticos

Sumado a lo ya presentado, la imposición de funciones de estatus en el término ‘Y’ de la regla constitutiva de un hecho institucional impone, a su vez, deberes y obligaciones en la realidad institucional. Tal y como señala Searle:

[...] la gran división en la categorización de la realidad institucional es entre lo que el agente *puede* hacer y lo que el agente *debe* (y no debe) hacer, [en específico], entre lo que el agente está *habilitado* para hacer y lo que a él o ella se le *requiere* hacer como resultado de una asignación de estatus especificada en el término Y. (1995, p. 100-101)

De este modo, por ejemplo, si consideramos el hecho institucional ‘los psiquiatras cuentan como las únicas personas calificadas para determinar una disforia de género en Chile’, la asignación de función estipulada al objeto ‘psiquiatra’ a partir de lo especificado en el término ‘Y’ implicará determinados deberes en cuanto a la función. En este caso en particular, un deber asociado a la función de estatus descrita es declaración de la disforia de género en caso de identificarla. Un psiquiatra no debe guardarse para sí su diagnóstico, independiente de que, por razones no médicas y siguiendo sólo su criterio moral, es lo mejor para su paciente (tal cual como sería en el caso de un psiquiatra fanático religioso que considera que la disforia de género no es un tema que tratar por motivos de fe en las escrituras).

Otro ejemplo de un deber vinculado con la práctica de la psiquiatría es que, una vez realizado el diagnóstico, es deber del psiquiatra realizar los procedimientos médicos y administrativos asociados, como asesorar en la toma de medicamentos a su paciente o escribir la receta que habilitará al

paciente luego a comprar los medicamentos indicados. Finalmente, otro deber que se desprende de esta asignación de función de estatus es que el psiquiatra debe evitar someter a su paciente a prácticas inmorales a favor de su diagnóstico. Por consiguiente, y atendiendo a los ejemplos esbozados, los poderes deónticos derivados de la asignación de funciones de estatus pueden ser formalizados, según Searle, de la siguiente manera: “Nosotros reconocemos colectivamente (Y existe en C porque (SRY (S tiene el poder (S hace A))” (2010, p. 102).

Dado que la asignación de funciones de estatus requiere de intencionalidad colectiva, nosotros reconocemos que nuestro término ‘Y’, a saber, ‘ser las únicas personas calificadas para determinar una disforia de género’ existen nuestro ‘C’, ‘Chile’ porque ‘S’, ‘los psiquiatras’ tienen una *relación* con nuestro término ‘Y’ en la cual ‘S’ no es sino *idéntico* al contenido de nuestro término ‘Y’. Por consiguiente, ‘los psiquiatras’ tienen el poder de ser aquellos que hagan ‘A’, donde el término ‘A’ puede ser cualquiera de las tres acciones constituyentes de deber que hemos señalado, a saber: i) ‘enunciar al paciente diagnóstico’; ii) ‘realizar los procedimientos médicos y administrativos asociados al diagnóstico’; iii) ‘evitar actos dañinos para el paciente al realizar su diagnóstico’.

Como hemos visto a lo largo de este apartado, la realidad institucional se compone de la asignación de funciones de estatus, de la intencionalidad colectiva y de las reglas constitutivas. Con esto hemos sentado una base teórica en torno a los fundamentos de la ontología social. Sin embargo, antes de abocarnos a nuestra tarea de vincular esta con la epistemología del testimonio, debemos atender a cómo ha sido caracterizado el testimonio en la literatura especializada. Procederé, por tanto, a dar cuenta de qué es el testimonio en la epistemología social.

Definiciones de testimonio

Luego de que Cecil Coady publicara su *Testimony: a philosophical study* en 1992, diversas definiciones ontológicas y epistemológicas del testimonio surgieron. Sin embargo, el modo de describir ontológicamente el testimonio se

encuentra directamente ligado con las determinaciones epistémicas que los epistemólogos atribuyen a este como fuente de justificación y/o conocimiento. Si bien la variedad de definiciones no ha sido escasa, quisiera centrarme en tres que resaltan a lo largo de la tradición: en primer lugar, la de Cecil Coady por ser la instauradora de la epistemología del testimonio como tal; la de Peter Graham, por ser una respuesta y modificación directa de la de Coady; y finalmente la de Jennifer Lackey, por romper con el esquema predominante que considera al testimonio como un fenómeno sólo desde la perspectiva del hablante. Procederé, por lo tanto, a analizar cada una de estas para dar cuenta de cómo ha sido caracterizado el testimonio ontológicamente.

Testimonio de acuerdo con Coady

Cecil Coady señala que un testimonio natural³ consiste en que un hablante S testifica por medio de una afirmación *p* si, y sólo si:

Ct₁) Su afirmación que *p* es evidencia de *p* y es ofrecida como evidencia de *p*.

Ct₂) S posee las competencias relevantes, autoridad o credenciales para afirmar verdaderamente que *p*.

Ct₃) La afirmación de S que *p* es relevante para alguna disputa o cuestión sin resolver —la cual puede ser, o no ser, *p* en cuestión— y es dirigida a aquellos que requieren evidencia en la materia. (1992, p. 42)

Analizaremos esta postura a la luz del siguiente ejemplo: imaginemos que en una conversación casual entre amigos discutimos acerca de los

³ Coady opone la noción de testimonio natural a la noción de testimonio formal: por una parte, el primero refiere al acto de ‘relatar’ algo en la vida cotidiana, tal cual una persona le puede contar a otra una experiencia vivida; mientras que, por otra, el segundo refiere al acto de testificar en un contexto judicial. Si bien ambas nociones son similares estructuralmente, cabe señalar que todo testimonio formal posee una mayor complejidad en cuanto a su definición: posee tanto las tres cláusulas que constituyen la definición del testimonio natural junto a otras que el autor desarrolla a lo largo del capítulo. Cfr. Coady, 1992, p. 25-42.

posibles efectos negativos del consumo de cafeína para el organismo: por una parte, yo afirmo que esta realmente es dañina para el cuerpo humano y que no debe ser consumida; mientras que, por otra parte, un amigo (a_1) me señala lo contrario y avala la tesis de que el consumo de cafeína no implica ningún efecto dañino para la salud. Frente a tal disputa, un segundo amigo interfiere a favor de mi postura señalando que hay un estudio x que justifica mi juicio. Teniendo esto en consideración, ¿sería la afirmación de (a_2) un caso de testimonio? De acuerdo con el planteamiento de Coady no, pues si bien la intervención de (a_2) cumple con (Ct_1) en la medida de que su intervención pretende ser una evidencia y se ofrece como tal, y además cumple con (Ct_3) dado que pretende resolver el conflicto o discusión acerca de los posibles efectos negativos de la cafeína que sostenemos (a_1) y yo, no cumple con (Ct_2) dado que este, es decir, (a_2), no posee ninguna suerte de acreditación relevante en lo que concierne a conocimientos de dietética, por lo que su intervención no puede ser considerada, a la luz de la definición de Coady, un testimonio.

Ahora bien, distinto sería si fuese el caso que (a_2) posea un grado académico en nutrición o algún elemento análogo que pudiera dar cuenta de sus conocimientos. Si poseyera una acreditación que dé cuenta de sus conocimientos su intervención cumpliría cabalmente con (Ct_1), (Ct_2) y (Ct_3), por lo tanto, nos encontraríamos con un testimonio. Como hemos visto, esta definición implica que para que una afirmación constituye ontológicamente un testimonio, la intención del hablante de presentar su palabra como evidencia para resolver un conflicto en conjunto con la acreditación de conocimiento deben presentarse en conjunto. De esta manera, el testimonio queda supeditado ontológicamente no sólo a la emisión de una aseveración, sino también a la capacidad de demostrar que el sujeto posea de manera efectiva la acreditación correspondiente para poder testificar al respecto.

Testimonio de acuerdo con Graham

Considerando el problema de la acreditación en la definición ontológica del testimonio de Coady, Graham plantea una postura diferente: en efecto, señala que un hablante S testifica afirmando que p si, y sólo si:

Gt₁) La afirmación de S que *p* es ofrecida como evidencia de *p*

Gt₂) S pretende que su audiencia crea que posee la competencia relevante, autoridad o credencial para afirmar verdaderamente que *p*

Gt₃) La afirmación de S que *p* es considerada por S como relevante para algún problema que él cree es discutido o no se encuentra resuelto (el cual puede ser o no ser *p*) y es dirigido a aquellos que él cree requieren evidencia sobre el asunto. (Graham, 1997, p. 227)

Considerando la definición dada por Graham, podemos evidenciar que difiere de la de Coady, en primer lugar, en cuanto a la noción de evidencia encarnada en (Ct₁). Según Graham, la noción de evidencia que maneja Coady posee tres sentidos, es decir, en tanto que *evidencia potencial*, en tanto que *evidencia verídica* o como *evidencia de x*: en primer lugar, la evidencia potencial, según el mismo Graham, “requiere de una conexión objetiva, asociación o regularidad entre la pieza de evidencia putativa y aquello para lo que es evidencia” (1997, p. 227). No obstante, puede ser el caso que esta no resulte ser verdadera; por otra parte, la evidencia verídica corresponde a una evidencia potencial que sí resulta ser verdadera; finalmente, la *evidencia de x* corresponde a aquello que S toma por evidencia de que *x* es el caso. Esta distinción es de suma importancia puesto que, como pretende demostrar Graham con su caso del ‘millonario’, la definición de Coady del testimonio y el concepto de evidencia encarnado en ella son demasiado restrictivos. Consideremos el siguiente el caso de Graham para dar cuenta de lo dicho:

Un millonario, un ser humano adulto normal, se encuentra varado junto con una estrella de cine, un profesor y otras personas en una isla remota. Un día, un barril con botellas de vino sin etiqueta es arrastrado [por la marea hacia la isla]. Sólo el millonario sabía realmente algo sobre vinos. Él afirma que, por medio de la degustación, puede distinguir los diferentes vinos. Desafortunadamente y sin saberlo, ha perdido su paladar discriminatorio. Cuando el habla de los vinos al resto, ellos aceptan como verdadero lo que dice, aunque lo que él diga no sea mejor que la casualidad. [Por lo tanto,] sus declaraciones sobre los vinos no son

evidencia potencial acerca de los vinos, a pesar de que él y su público crean que él habla verdaderamente sobre ellos. (1997, p. 228)

En el caso expuesto, el millonario de Graham pierde la conexión objetiva entre su paladar discriminatorio y aquello para lo que es evidencia, esto es, ser capaz de diferenciar los diferentes vinos independientes de su etiqueta. Sin embargo, como queda señalado en el ejemplo, el millonario sigue siendo capaz de testificar que *p* y sus interlocutores creen su testimonio como verdadero independiente de que su identificación de los diferentes tipos de vinos no sea más que una apuesta a ciegas sobre ellos. Por lo tanto, de acuerdo con Graham, en comparación con (Gt₁), (Ct₁) de Coady no es una cláusula adecuada para definir lo que es un testimonio. Asimismo, en consecuencia, tanto (Ct₂) como (Ct₃) tampoco lo son adecuadas, pues el millonario ha perdido su ‘credencial’ como degustador de vinos a la par que ha perdido su facultad para diferenciarlo. Además, en la situación descrita no hay ninguna exigencia de evidencia por parte de los interlocutores, sino más es el mismo millonario quien considera relevante testificar al respecto —sea por los motivos que sea—. Por lo que (Gt₂) y (Gt₃) constituyen una mejor alternativa a la postura ‘estrecha’ expresadas en las condiciones de Coady para la existencia del testimonio.

Diferencia de roles

Ahora bien, a pesar de que las condiciones de Graham suscitan, a primeras, una opción viable, de acuerdo con Lackey, esta como la de Coady, “fallan en reconocer la distinción entre *el testimonio como un acto intencional del hablante y el testimonio como una fuente de creencia o conocimiento para el oyente*” (2008, p. 27), esto es, no identifican la diferencia de roles que suscita como fenómeno social. Atendiendo a esta problemática, la misma Lackey define el testimonio señalando que, por una parte, S produce un *testimonio del hablante* acerca de *p*, si y sólo si: “(TH) al realizar un acto comunicativo *a*, S razonablemente tiene la intención de transmitir información acerca de *p* (en parte) en virtud del contenido comunicable de *a*” (2008, p. 30).

Asimismo, por otra parte, señala que S testifica para el oyente, desde la perspectiva del oyente si y sólo si: “(TO) al realizar un acto comunicativo *a*, el oyente de S razonablemente asume que el acto transmite la información que *p* (en parte) en virtud del contenido comunicable de *a*” (2008, p. 32).

De acuerdo con dicha distinción, el hablante posee el rol social de ser portador, por medio de su testimonio que *p*, de un contenido comunicable que es producido de forma intencionada para ser transmitido a un oyente; mientras que, por otra parte, el oyente posee el rol social de participar de la comunicación como receptor de la información comprendida en el contenido comunicable emitido intencionalmente por el hablante. De este modo, el testimonio se constituye en esta definición como un fenómeno relacional de intercambio de conocimiento en el que los roles de los participantes de este difieren en cuanto a los roles que desempeñan en cuanto al fenómeno.

En el caso de las definiciones de Coady y Graham, el testimonio era definido sólo en torno a la postura del hablante, sin considerar la naturaleza social del testimonio. Sin embargo, la definición de Lackey es integradora y no depende de la noción de evidencia externa ni de credencial, como lo sería en un reduccionismo ontológico (tal como sucede en el caso de Coady y, *mutatis mutandis*, en el caso de Graham), sino que simplemente estipula que el testimonio constituye un acto comunicativo.

Ahora bien, habiendo descrito cómo operan las definiciones de Coady, Graham y Lackey debemos hacernos las siguientes preguntas: ¿alguna es compatible con la ontología social de John Searle? ¿es posible incrustar alguna de estas en el aparato conceptual de la realidad institucional? A primeras, debemos descartar las definiciones de Coady y Graham para esta tarea, pues sus definiciones carecen de dimensión social, por lo que sólo nos queda la definición de Lackey. Sin embargo, esta, si bien da cuenta de aspectos sociales del testimonio, aún no es suficiente para pensar el testimonio dentro de los parámetros de la realidad institucional pues no nos permite divisar ningún elemento por el cual supeditar la epistemología del testimonio a la propuesta searleana. Por lo tanto, procederé en el siguiente apartado a dar cuenta cómo puede ser caracterizado el testimonio dentro de la ontología

social de Searle y qué elementos podemos considerar de la epistemología del testimonio para esta labor.

El testimonio como hecho institucional

Si queremos explotar la posibilidad de incrustar la epistemología del testimonio dentro de la ontología social de Searle, deberemos seguir un camino distinto. Si bien las definiciones dadas por la epistemología social no son adecuadas para esta labor, sí hay propuestas que nos pueden servir para esto. Una de estas es el denominado enfoque interpersonal del testimonio, la cual concibe al testimonio como un acto de habla asertivo. ¿Cómo se relaciona esto con la ontología social de Searle? Procederé, en lo que resta del trabajo, a caracterizar el testimonio como un hecho social y dar cuenta de cómo en enfoque interpersonal del testimonio nos puede servir para delimitar los poderes deónticos asociados a este.

Estructura institucional del testimonio

El testimonio es una institución caracterizable por medio del aparato conceptual de la ontología social de Searle. Dado que constituye una institución, este deberá instaurarse por medio de una regla constitutiva derivada de la asignación de una función de estatus aceptada por una comunidad. Dicha *regla constitutiva del testimonio* puede ser caracterizada de la siguiente manera:

RCT) ‘La palabra de S’ cuenta como ‘evidencia fundada epistémicamente en sí misma para el intercambio de conocimiento’ en ‘circunstancias en la que se presenta *dependencia epistémica*⁴’.

⁴ El concepto de dependencia epistémica fue acuñado por John Hardwig para señalar que, dado que no podemos justificar todas nuestras creencias, dependemos de los expertos para *conocer* aquello de lo cual somos legos. Asimismo, este concepto fue profundizado por Elizabeth Fricker señalando que no sólo hay dependencia entre expertos/legos, sino también en cuanto al ‘mejor posicionamiento’, esto es, frente a aquellos que pudieron experimentar un hecho al cual nosotros no podemos acceder. Cfr. Fricker, 2006; Hardwig, 1985.

Como se ve en la regla, se ha asignado la función ‘evidencia fundada epistémicamente en sí misma para el intercambio de conocimiento’ al objeto ‘La palabra de S’ en el contexto general ‘circunstancias en las que se presenta dependencia epistémica’. Siendo esto así, ¿por qué esta regla es constitutiva y no más bien regulativa? Se podría afirmar que el testificar es una actividad previa a la regla y que esta sólo puede normarlo de forma regulativa. Sin embargo, el testimonio sólo puede manifestarse como fenómeno social allí donde la palabra de un sujeto en específico es considerada como la evidencia del contenido proposicional que está siendo aseverado por medio de la realización de un acto de habla asertivo. De este modo, por ejemplo, si yo testifico a mi hermano que, caminando por la calle, presencié un asesinato, el acto mismo de testificar dependerá, para lograr su cometido (esto es, el intercambio de conocimiento del hablante al oyente), que de mi hermano pueda considerar mi palabra como evidencia de tal hecho, de modo que, por consiguiente, crea en mí y en la verdad del contenido aseverado.

Este reconocimiento se supedita a la intencionalidad colectiva. Es colectivamente aceptado, al menos en Chile, que la palabra de las personas en cuanto a testimonios naturales cuenta como evidencia de ello. Por ejemplo, si voy en la calle y le solicité a un caminante que me informe qué hora es en ese momento porque requiero de dicha información, al escuchar su testimonio respecto de la hora (suponiendo que no alcanzo a ver su reloj o celular) no pondré si quiera en cuestión su palabra y lo asumiré como verídico en la medida de que constituye la única evidencia necesaria para la formación de estados doxásticos epistémicamente justificados en torno a la información entregada en el intercambio testimonial.

Ahora bien, el testimonio, me parece, constituye un caso especial de hecho institucional. Al igual que sucede con el lenguaje, este posibilita otros hechos institucionales. Searle dice que el lenguaje posee primacía ontológica sobre las otras instituciones en la medida de que esta posibilita la realidad social: en efecto, señala que:

[...] la capacidad de anexar sentido, una función simbólica, a un objeto que no tiene dicho sentido intrínsecamente es la precondition no sólo

del lenguaje sino de toda la realidad social. La capacidad preinstitucional para simbolizar es la condición de posibilidad de la creación de todas las instituciones humanas. (Searle, 1997, p. 75)

¿Posee el testimonio una capacidad similar? El testimonio es, en última instancia, confianza, y sólo en la medida de que somos capaz de confiar en la palabra del otro, en que sus enunciados asertivos no son malintencionados o no son, lisa y llanamente, mentiras, podemos establecer el requisito de la intencionalidad colectiva en la construcción de la realidad institucional.

Dado que sólo hay intencionalidad colectiva donde hay acuerdo declarado, la confiabilidad en torno a las palabras de los demás respecto de ámbitos de conocimiento, es decir, el testimonio, constituye una pieza fundamental en la asignación de funciones de estatus. Me explico: imaginemos una sociedad en la que nadie recibe testimonio ajeno pues esta suscita una institución confiable, pues es sabido que todos mienten patológicamente. Ante una situación así, ¿es posible llegar a alguna clase de acuerdo? Y si fuese posible, ¿serían realmente acuerdos si estos se basan en mentiras? Si no es posible el acuerdo, tampoco la intencionalidad colectiva; y si no es posible la intencionalidad colectiva, tampoco lo es la imposición de reglas constitutivas a partir de la asignación de funciones de estatus aceptadas colectivamente. En consecuencia, se imposibilita la realidad institucional. De este modo, el testimonio como institución no puede ser desestimada y se presenta, en tanto que confianza en la palabra del otro, como un factor preintencional necesario para los postulados de la ontología social de Searle.

Garantía, honestidad y exigencia de credibilidad como poderes deónticos

Siguiendo lo dicho, ahora debemos atender a los poderes deónticos derivados del testimonio como institución. Para esto, quisiera, en primer lugar, hablar de los postulados de Richard Moran en torno al enfoque interpersonal del testimonio. Según Moran, al testificar:

[...] hablante y audiencia no confrontan la preferencia como si esta fuese un fenómeno con un estatus epistémico independiente o natural que puedan evaluar como tal, pues el hablante no confronta en absoluto su propia aserción como si se tratase de un fenómeno, sino más bien como un asunto relativo a su compromiso. (Moran, 2006, p. 294)

Asimismo, la consideración del hablante de su testimonio como un compromiso confiere al oyente la capacidad de juzgar al hablante pues, en tanto que este da su palabra, se implica a sí mismo como una garantía (*assurance*) de su afirmación. Por lo tanto:

Cuando todo va bien en el testimonio, el hablante da a su audiencia una razón para creer algo, pero, a diferencia de otras maneras de inducir creencias en otros, en este caso la razón que le es provista a la audiencia es vista, por ambas partes, como dependiente de que el hablante se haga responsable de ello, confiriendo un derecho a queja para su audiencia en caso de que su afirmación sea falsa. (2006, p. 295)

En otras palabras, la posición del hablante suscita servir como garante y fuente epistémica de la expresión de un contenido comunicable para una audiencia que, en virtud de su propio rol, es capaz de exigir que la información conferida en el testimonio de S respecto de *p* sea, en efecto, verídica, considerando con ello la relación de confianza como una suerte de contrato que, al no cumplirse cabalmente, es susceptible de ser denunciado y juzgado como una falta de cumplimiento a las base de la relación entre el hablante y el oyente.

Siendo esto así, ¿cómo es que estas consideraciones epistémicas sobre el testimonio se relacionan con los poderes deónticos asociados a este en la realidad institucional? ¿Cuáles son los poderes deónticos asociados a la práctica testimonial? Identificó fundamentalmente tres: la garantía, la honestidad y la exigencia de credibilidad.

Garantía como poder deóntico

Siguiendo lo dicho por Moran, el testimonio no simplemente asevera el contenido de la proposición en cuestión, sino que también suscita un compromiso con la verdad de este. Esto debido a que el testimonio, además de ser un acto de habla asertivo, es también un acto de habla compromisorio. Cuando testifico que *p* no digo simplemente '*p* es verdadero' sino también, por medio de dicha enunciación, implíco tácitamente la promesa 'prometo que lo que digo es verdadero', constituyéndose así como un acto de habla indirecto. Por consiguiente, si el testimonio contiene promesas respecto del contenido aseverado, esto implicará que S posee un deber asociado a esta práctica, esto es, presentarse como garantía de lo aseverado al testificar. Por lo tanto, siguiendo el esquema formal de Searle, la *garantía como poder deóntico* se representa de la siguiente manera:

GPD) Nosotros reconocemos colectivamente (contar como 'evidencia fundada epistémicamente en sí misma para el intercambio de conocimiento' existe en 'circunstancias en la que se presenta *dependencia epistémica*' porque (SRY (S tiene el poder (S se presenta como garantía de lo aseverado al testificar)

De este modo, la garantía constituye un deber asociado al testimonio como institución el cual debe cumplirse al testificar para una correcta realización del intercambio testimonial.

Honestidad como poder deóntico

Este segundo poder deóntico deriva del anterior pues, comprometerse con la veracidad de lo aseverado también suscita que lo aseverado y lo prometido hayan sido realizados honestamente. Al testificar, debo hacerlo siempre con honestidad, pues de lo contrario, no hay testimonio realmente. En consecuencia, siguiendo el esquema formal de Searle, la *honestidad como poder deóntico* se representa de la siguiente manera:

HPD) Nosotros reconocemos colectivamente (contar como 'evidencia fundada epistémicamente en sí misma para el intercambio

de conocimiento' existe en 'circunstancias en la que se presenta *dependencia epistémica*' porque (SRY (S tiene el poder (S asevera honestamente al testificar)

De este modo, la honestidad se configura como un deber más asociado a la institución del testimonio. Este es, quizás, el deber más importante para el intercambio testimonial. Como dijimos anteriormente, una sociedad de mentiroso patológico no puede mantenerse. Sin embargo, aunque no sean todos, si hay mentirosos patológicos en nuestra sociedad. Estos no respetan este deber por lo que, en estricto rigor, no testifican. No obstante, no deja de ser una preocupación que puedan simular honestidad y colar en las personas, debido precisamente a la confianza a priori en el testimonio, creencias falsas o malintencionadas.

Exigencia de reevaluación como poder deóntico

Este poder deóntico es más bien asociado al oyente y deriva de (a). En la medida de que se hace una promesa al testificar, el hablante le cede un poder deóntico al oyente, a saber, la exigencia reevaluación del estatuto epistémico de la creencia. Según Searle, “[...] si te hago una promesa, entonces tú, en efecto, tienes un poder deóntico hacia mí, porque he creado un motivo que me enlaza a mí con lo que prometí hacer” (Searle, 2010, p. 148). Ese vínculo recae en el oyente que puede exigir al hablante, en caso de equivocarse al testificar, que este reevalúe el estatuto epistémico de sus propias creencias. Esto es lo que se ha llamado en la, literatura especializada, ‘humildad intelectual’ de acuerdo con el enfoque doxástico⁵. Esta puede ser hecha o no por el propio sujeto o puede ser exigida externamente. Sin embargo, para efectos de nuestra exposición, la exigencia será externa pero el sujeto será internamente responsivo a esta exigencia. Por lo tanto, siguiendo el esquema formal de Searle, la *exigencia de reevaluación como poder deóntico* se representa de la siguiente manera:

⁵ Cfr. Church y Samuelson, 2017, p. 7.

ERPD) Nosotros reconocemos colectivamente (contar como ‘evidencia fundada epistémicamente en sí misma para el intercambio de conocimiento’ existe en ‘circunstancias en la que se presenta *dependencia epistémica*’ porque (SRY (S tiene el poder (S responde a las exigencias de reevaluación)

Hemos de notar, antes de terminar sobre este punto, que las exigencias de reevaluación no implican necesariamente malicia de parte del hablante. Bien puede ser el caso que el hablante se equivoque en la información y que esta no sea, en estricto rigor, su responsabilidad. Por ejemplo, podría haber sido timado, seleccionar una fuente confiable de conocimiento o simplemente haber llevado a cabo, involuntariamente, una modificación del contenido que suscita posteriormente que el testimonio sea falso. Lo importante a destacar aquí es que el estatuto epistémico del hablante al testificar puede ser cuestionado y es deber de este responder de manera adecuada a tal cuestionamiento.

Así, aplicando lo mencionado en esta sección, podemos analizar el siguiente caso: Juan asiste a una consulta de rutina y su dermatólogo, Marco, le diagnostica un cuadro de dermatitis seborreica. Juan, que es mecánico, depende de Marco para saber qué afección tiene; al fin y al cabo, él sabe de vehículos y no de la estructura de la piel, por lo que, aunque no tenga ningún impedimento intelectual para saber de medicina, no tiene los recursos (tiempo, energía, entre otros) para estudiar esa carrera si se dedica por completo a su profesión. Juan entonces le cree a Marco, pues Marco tiene el poder deóntico (es decir, una habilitación y un deber) de diagnosticarle presentándose a sí mismo como una garantía de ese diagnóstico por su acreditación profesional. Asimismo, tiene el poder deóntico de ser honesto, de modo que era garantía no es meramente trivial, sino que va de la mano con una virtud epistémica conducente a la verdad. Yendo todo bien, Juan tendría una solución a su problema.

Ahora bien, ¿qué pasaría si Marco, el dermatólogo, se equivoca? ¿Juan tiene un derecho epistémico a cuestionar el diagnóstico si ve que no le funciona? ¿Tiene algo que reclamarle a Marco? Juan, en su condición de

dependencia epistémica de lego-experto, tiene toda la potestad de alegarle a Marco por su equivocación. Sin embargo, ello no es un problema mayor: ya que Marco es una autoridad epistémica, ha de cumplir con el poder deóntico de la exigencia de reevaluación. Dado que ningún diagnóstico da cuenta con necesidad de lo que sucede con el individuo (es decir, se basan en observaciones con evidencia disponible limitada), bien que puede luego eso ajustarse con una reevaluación a la luz de la nueva evidencia, a saber, que el tratamiento dado no haya funcionado. Yendo todo bien, Marco calibraría su testimonio médico con el nuevo antecedente y con cómo este es responsivo con su trasfondo de creencias relativas a la pericia médica, logrando dar con una solución para Juan (y así cuantas veces sea necesario).

Este punto es de suma importancia, pues un pseudoexperto difícilmente testificaría siendo susceptible de reevaluar genuinamente su testimonio. En el caso de un homeópata que pretende dar diagnóstico médico sin tener pericia al respecto, ante la imposibilidad de atender correctamente las necesidades de sus pacientes, quizás ajuste el supuesto tratamiento. Sin embargo, difícilmente podemos decir que eso se haría presentándose como una garantía (generalmente no se hacen responsables), siendo honestos (aceptando que no saben lo que dicen saber) y reevaluando sus creencias (aceptando que están en un error). Aquel que integra, no obstante, estos poderes deónticos propios de la práctica testimonial, sí puede decir que ha dado testimonio genuino al respecto y no estas formas parasitarias de pseudopericia que, lamentablemente, nos podemos encontrar en nuestra interacción con otros.

Conclusiones

Para finalizar, podemos concluir que la posibilidad de incrustar la epistemología social en la propuesta ontológica de Searle es una vía de trabajo factible y, probablemente, muy enriquecedora para ambas partes. Como hemos visto, es posible interpretar el testimonio como una institución con poderes deónticos asociados. Esta interpretación nos ha dado una mayor precisión de qué es el testimonio ontológicamente a diferencia de las definiciones, bastante aisladas de la ontología social, que han dado los

epistemólogos del testimonio a lo largo de décadas. Muchos puntos quedan por precisar respecto de la posibilidad de esta labor, y varias de las ideas que se han quedado en el intento son de interés sistemático para dicha tarea. Ejemplo de esto es, por una parte, las implicancias normativas implícitas no intencionales que suscita nuestra experiencia del testimonio; por otra parte, si el testimonio es una institución fundamental, cuál es precisamente su posición en torno al lenguaje y al resto de los hechos institucionales, sobre todo si consideramos los aportes que la etología cognitiva podría otorgarnos sobre el paso del comportamiento colectivo a la instauración de la verdad y la confianza en la naturaleza (Tomasello, 2009); finalmente, también de suma importancia, no hemos podido, por la extensión de este trabajo, explorar el asunto más importante de esta pretensión de vinculación sistemática, a saber, cómo es que la realidad institucional influye en las determinaciones epistémicas del hablante y el oyente en el intercambio testimonial, en específico, qué nivel de causación suscita la ontología social respecto de las disposiciones epistémicas de los individuos y cómo la epistemología puede elaborar parámetros normativos que constriñan cómo ha de ser llevado el intercambio testimonial. Estas preguntas y, probablemente muchas más, quedarán pendientes.

Agradecimientos

Fuente de financiamiento: "ANID-PFCHA/Doctorado Nacional/Año 2022 —21220627)

Referencias

- Church, I. M., & Samuelson, P. L. (2017). *Intellectual humility: An introduction to the philosophy and science*. Bloomsbury.
- Coady, C. A. J. (1992). *Testimony: A philosophical study*. Oxford University Press.
- Fricker, E. (2006). Testimony and epistemic autonomy. En J. Lackey & E. Sosa (Eds.), *The epistemology of testimony* (pp. 193–224). Oxford University Press.
- Graham, P. J. (1997). What is testimony? *The Philosophical Quarterly*, 47(187), 227–232.
- Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7), 335–349.
- Lackey, J. (2010). The nature of testimony. En J. Lackey, *Learning from words: Testimony as a source of knowledge* (pp. 26–52). Oxford University Press.

- Moran, R. (2006). Getting told and being believed. En J. Lackey & E. Sosa (Eds.), *The epistemology of testimony* (pp. 272–306). Oxford University Press.
- Searle, J. R. (1995). *The construction of social reality*. Penguin Books.
- Searle, J. R. (1998). *Mind, language, and society: Philosophy in the real world*. Basic Books.
- Searle, J. R. (2010). *Making the social world: The structure of human civilization*. Oxford University Press.
- Smith, B. (2010). John Searle: From speech acts to social reality. En B. Smith (Ed.), *John Searle* (pp. 1–33). Cambridge University Press.
- Tomasello, M. (2009). *Why we cooperate*. The MIT Press.

Antropología filosófica de la memoria: Antecedentes desde Lluís Duch a la formación cultural del *anthropos**

Philosophical anthropology of memory: Background from Lluís Duch to the cultural formation of the *anthropos*

Carolina Lagos Oróstica

clagos@ucsc.cl

Universidad Católica de la Santísima Concepción

ORCID ID 0000-0001-7858-2985

Resumen: El presente artículo de reflexión tiene por objetivo exponer una interpretación sobre la relevancia de la memoria en la formación del ser humano como ser cultural. De acuerdo con los postulados filosóficos-antropológicos del monje benedictino Lluís Duch, el artículo se inicia situando a la época posmoderna como un periodo que obstaculiza el desarrollo de las capacidades humanas de vinculación y pertenencia cultural. En gran medida, esta situación se debe a una crisis pedagógica generada por las condiciones de la globalización y el capitalismo que se extienden a los modelos educativos, generando conocimientos meramente informacionales que no contribuyen a la formación integral de la persona. Para Lluís Duch, la trayectoria vital del ser humano, es decir, todas sus actuaciones reflexivas y elaboraciones prácticas, se insertan en un marco cultural que necesariamente debe acudir al aprendizaje. No obstante, en la actualidad los seres humanos se encuentran

* Artículo de reflexión derivado de la investigación “*La estructura logo-mítica humana: una comprensión de la religiosidad desde la antropología filosófica de Lluís Duch*”, realizado por la autora mediante financiamiento otorgado por la Vicerrectoría de Investigación y Posgrado de la Universidad Católica de la Santísima Concepción, Concepción, Chile.

ajenos a formas educativas mediadoras y trasmisoras de saberes significativos que permiten el despliegue de sus capacidades *cosmizadoras*; dadoras de sentido, orden y unidad. Lo anterior, afecta directamente a la formación del ser humano como ser cultural. Al respecto, este ensayo se compone de dos partes. La primera, caracteriza la esfera educativa en la sociedad posmoderna, distinguiendo la incapacidad de los sistemas educativos actuales en su función transmisora de saberes para la formación de la dimensión cultural humana. La segunda, se refiere a la relevancia de la memoria como factor imprescindible para la formación del sujeto cultural. Acorde con el estudio hermenéutico efectuado por Lluís Duch, el ensayo es elaborado con un método descriptivo e interpretativo con la finalidad de hallar el sentido de la relación entre educación, cultura y formación humana.

Palabras clave: Cultura; educación; memoria; persona; posmodernidad.

Abstract: This reflective article aims to present an interpretation of the relevance of memory in the formation of human beings as cultural beings. In accordance with the philosophical-anthropological postulates of the Benedictine monk Lluís Duch, the article begins by situating the postmodern era as a period that hinders the development of human capacities for cultural connection and belonging. To a large extent, this situation is due to a pedagogical crisis generated by the conditions of globalization and capitalism that extend to educational models, generating purely informational knowledge that does not contribute to the integral formation of the person. For Lluís Duch, the human life trajectory, that is, all its reflective actions and practical elaborations, is inserted into a cultural framework that necessarily must involve learning. However, human beings today are alienated from educational forms that mediate and transmit meaningful knowledge that allows them to develop their *cosmizing* capacities, which give meaning, order and unity. This directly affects the formation of human beings as cultural beings. In this regard, this essay consists of two parts. The first characterizes the educational sphere in postmodern society, distinguishing the inability of current educational systems to transmit knowledge for the formation of the human cultural dimension. The second part refers to the relevance of memory as an essential factor in the formation of the cultural subject. In accordance with the hermeneutic study carried out by Lluís Duch, the essay is written using a descriptive and interpretative method with the aim of finding the meaning of the relationship between education, culture and human formation.

Keywords: Culture; education; memory; person; postmodernity.

Introducción

El análisis que el filósofo, teólogo y antropólogo catalán Luis Duch Álvarez (1936-2018) realiza sobre el ser humano y la cultura, comprende ámbitos que abarcan “la familia, la ciudad, la vida cotidiana, la comunicación, las transmisiones, la salud, el mito, el símbolo, el rito” (Moreta, 2019, p. 9). La amplitud del examen expuesto en la obra del monje de Monserrat, constituye un profuso aporte filosófico-antropológico que destaca su vocación como pensador de *lo humano* (Duch, 2010). En este sentido, la hermenéutica sobre la situación existencial y cultural del *anthropos* desarrollada por Lluís Duch, muestra que su intención no consiste en realizar una descripción aislada de las particularidades de la especie humana, más bien, el propósito que mantiene es contribuir al conocimiento humanista sobre el ser humano y sus potencialidades obligatoriamente desenvueltas en el seno de la cultura y la sociedad. De esta manera, Duch traza un reconocimiento de la integralidad de la especie humana mediante una metodología hermenéutica, capacitada para descubrir desde sus diferentes dimensiones el sentido signifiante del «factor biográfico» del ser humano como ser estructuralmente *ambiguo* (Moreta, 2019).

Duch sostiene que la trayectoria vital del *anthropos* nunca deja de estar vinculada con la comunidad, la vida cotidiana y la cultura. Estos sectores resultan de un camino existencial intervenido por variados procesos positivos y negativos, efectuados por la especie humana con la intención de adquirir una estructuración y una contextualización histórica del *yo* y del *nosotros* (Duch, 2010). El autor agrega que todas las actuaciones reflexivas y todas las elaboraciones prácticas efectuadas por mujeres y hombres de toda época, se configuran como actividades que se insertan en el “marco y el campo de la cultura y la sociedad” (Duch, 2020, p. 81-82). Por consiguiente, las culturas pueden considerarse como específicas concreciones de las diversas actividades y del pensamiento de la especie humana (Duch, 2019). Es imposible imaginar que la persona pueda crecer y desarrollarse plenamente en un plano extracultural (Duch, 2010, 2020; Moreta, 2019). Como sostiene

Joan Carles Mèlich (2008): “Ser animales culturales, habitar culturas, es nuestro específico modo de vida, nuestra «profesión biológica». Hemos nacido casualmente en un tejido de historias, en un tiempo y en un espacio” (p. 103). Asimismo, la cultura es un fenómeno que surge debido a la condición relacional de una especie capacitada para poner en común (Choza, 2013) variadas tramas de significaciones (Geertz, 2003), que tienen por objetivo organizar “el espacio y el tiempo para que se vuelvan susceptibles de ser *habitados*” (Duch, 2012a, p. 52).

En el plano de la cultura el aprendizaje es fundamental, y “puede ser considerado como la forma más específica y auténtica de vivir como hombre” (Duch, 1997, p. 87). Lluís Duch examina que los sistemas de enseñanza en la sociedad posmoderna han adquirido un carácter mecánico puramente informacional. En la actualidad y de manera global, los seres humanos se encuentran ajenos a formas educativas mediadoras y trasmisoras de saberes significativos favorables para el despliegue de sus capacidades *cosmizadoras*; dadoras de sentido, orden y unidad. Más bien, los procesos formales de enseñanza hoy se ajustan a la situación de la segmentación y la autonomización de la vida (Solares, 2008). La posmodernidad asociada con “ideas de ruptura, cultura de lo nuevo, la presencia de nuevas tecnologías, globalización cultural, movilidad, aceleración en la historia, discontinuidad” (Jiménez, 2023, p. 60), experimenta una problemática situación en relación a la formación de la persona como ser cultural, puesto que existen serias deficiencias sociales para la integración de la persona en una realidad significativa y reconocida como un fondo de estabilidad en la que ella pueda “orientarse con una cierta autonomía y corrección por los senderos de la vida” (Duch, 2012a, p. 122).

Lluís Duch considera que la incompleta formación integral del ser humano, se manifiesta en el debilitamiento de su capacidad de interpretar, significar y *empalabrar* la realidad (Duch, 2019). Desde aquí, el autor propone ejercitar una pedagogía atenta a la metáfora, el símbolo, la narración y la memoria, porque el ser humano “a lo largo y ancho de su existencia no deja de operar con ellos y de vivir de ellos” (Duch, 2019, p. 19). Según Duch, el marco de referencia compuesto por la cultura no depende únicamente del

logos en su habilidad de conceptualizar las cosas, las realidades y el mundo. Las distintas culturas y la misma vida cotidiana, surgen de la plasticidad y la creatividad expresada en los universos simbólicos que son transmitidos por la memoria de los pueblos. En concordancia, la pedagogía a la que Duch aspira es una centrada en un humanismo renovado e interesado por la formación integral de la persona, con la finalidad de que ella consiga expresar “la verdad, la justicia y la libertad de lo humano” (Duch, 2019, p. 31). Es decir, el interés del autor es una educación formadora y orientadora del discernimiento que va más allá del funcionalismo y el pragmatismo. El propósito es superar la realidad de los modelos educativos de las sociedades posindustriales, habituadas a un lenguaje técnico-económico que no puede expresar adecuadamente “los procesos de humanización (especialmente en el campo cultural y pedagógico) de hombres y mujeres” (Duch, 2019, p. 31).

Por lo anterior, abordar la condición antropológica de la memoria y sus cualidades expresivas, es fundamental para el desarrollo de la persona como ser cultural. A este respecto, este artículo tiene por objetivo exponer una interpretación sobre la relevancia de la memoria en la formación del ser humano como ser cultural, considerando que la memoria favorece la transmisión de referencias culturales a través de tejidos narrativos que portan el pasado y la tradición. De acuerdo con la examinación sociocultural de Lluís Duch, el ensayo sitúa que la época posmoderna obstaculiza la actualización de las capacidades humanas relacionadas con la vinculación y pertenencia cultural. En gran medida, esta situación se debe a una crisis pedagógica generada por las condiciones de la globalización y el capitalismo. De acuerdo con esta premisa, en la primera parte del escrito, se expone una caracterización de la esfera educativa en la sociedad posmoderna. En la segunda, se describe interpretativamente el planteamiento filosófico-antropológico de Lluís Duch en relación a la pedagogía y la memoria en clave cultural. En su conjunto, el ensayo desea mostrar la posición humanista de Lluís Duch frente a los trayectos y fines de todas las creaciones humanas (Chillón, 2011). Desde allí, la intencionalidad reflexiva e interpretativa de este trabajo es elaborada a través de una hermenéutica que pretende ser acorde con el enfoque fenomenológico de Lluís Duch. Asimismo, la vía

hermeneútica favorece el ingreso a los significados que aportan sentido (Rodríguez, Gil & García, 1996) sobre el vínculo entre educación, cultura y formación humana.

Educación posmoderna

El diseño de la cultura en el mundo globalizado se realiza cada vez más distanciado de las condiciones que favorecen a la persona en la identificación de espacios y lugares originarios de su cultura. Actualmente, los vínculos y las identidades colectivas se desdibujan, como también los distintos saberes manuales o intelectuales (Chillón & Duch, 2016) son profundamente afectados por la homogenización de la cultura. No obstante, Duch observa que la cultura es una realidad indestructible e insuperable del ser humano, que “se manifiesta en todos los ámbitos de su pensamiento, acción y sentimientos” (Duch, 2015, p. 44). La cultura se expresa en una comunidad que, a través de su vida cotidiana, responde a una determinada “*tradición* de vida, de acción, y de sentimientos más o menos compartidos” (Duch, 2012a, p. 113). En este sentido, la preocupación del autor con respecto a la calidad de los sistemas pedagógicos, de las relaciones familiares, de los medios que transmiten los mensajes simbólicos y las referencias conceptuales, se identifica como una problemática que debe examinarse a la luz de la influencia de las visiones fragmentadas de la realidad y, al mismo tiempo, teniendo en cuenta las nociones mecanicistas y unilaterales que sobre el ser humano fomenta la posmodernidad (Duch, 2019).

Para Duch, la naturaleza cultural de hombres y mujeres (Duch & Mèlich, 2012) requiere ser comprendida en un sentido propiamente humanista; por consiguiente, los modelos educativos hoy tendrían el desafío de contribuir a la comprensión antropológica en ese mismo sentido. De acuerdo con el autor, concebimos que por sobre los objetivos de un conocimiento eficiente ordenado por el pago y la inversión (Lyotard, 1987), y más allá de los funcionalismos institucionales (Solares, 2003) que imponen metodologías de asesoramiento para la persona en formación, la pedagogía debe mantener su finalidad orientadora de las capacidades reflexivas del

anthropos, destinadas a descubrir las potencialidades de su entendimiento y su proceder.

Al mismo tiempo, desde la perspectiva sostenida por Duch, se puede argumentar que la pedagogía posmoderna ha perdido su ocupación de colaborar en la integración e identificación del ser humano con su comunidad. En efecto, el autor catalán expresa que la función de transmitir nociones para el habitar significativamente una realidad cultivada y compartida con sentido de lugar y pertenencia, es una situación por poco inaccesible para las personas de la posmodernidad. Este fenómeno global involucra a los modelos de enseñanza que hoy responden a la hegemonía del unidimensional *homo oeconomicus*, que concibe la mercantilización de todos los ámbitos sociales y personales (Chillón & Duch, 2016).

La impresión de Duch sobre los modelos de enseñanza es que carecen de fundamentos para educar la autocomprensión individual y la ubicación de la persona en una colectividad que debe acogerla y reconocerla (Duch, 1997). Por lo tanto, el autorreconocimiento de ser humano como integrante de una cultura que se sostiene en una cosmovisión axiológica, es un factor actualmente mermado, precisamente, por la escasez de referencias con respecto a qué significa *ser* seres humanos que viven en comunidad, coimplicados con procesos de transmisión y actuaciones testimoniales. En esta línea, “una «buena pedagogía» sería aquella capaz de «administrar adecuadamente» las presencias y las ausencias o las ausencias en las presencias, el pasado en el presente y el futuro en el presente” (Mèlich, 2008, p. 108).

Agrega Duch que la educación y la pedagogía deben contribuir con la *veracidad* y la *confianza* imprescindibles para el desarrollo sociocultural. “De este modo, el ser humano, puede constituirse como tal, es decir, pasar de las potencialidades inscritas en su mero instinto animal al campo de la artificiosidad, esto es, a la cultura o, mejor dicho, a *una* cultura concreta” (Duch, 2020, p. 104). Esta situación esencial, piensa el autor, hoy se encuentra en riesgo, debido a que la posmodernidad incrementa sus nociones de inestabilidad y relativismo mediante sistemas educativos que no actúan como

mediadores de referentes para las decisiones primordiales de la existencia humana (Duch, 2012a).

Centrados principalmente en la profesionalización del conocimiento y empleando reglas técnicas y utilitarias, los sistemas educativos se rinden “al rampante enfeudamiento y fragmentación de los conoceres” (Chillón, 2011, p. 36). La consecuencia es una experiencia educativa que no se desarrolla para “la formulación práctica de las grandes preguntas propias de la especie (...). Las que suelo llamar *preguntas fundacionales* por qué la vida, la muerte, el más allá, el mal, el amor, etc.” (Chillón, 2011, p. 281). Al mismo tiempo, la educación posmoderna no ha conseguido establecer ningún axioma para su propia reconducción como estructura que debe acoger y guiar al ser humano en su desarrollo. Por ello, Lluís Duch declara que es imperativo “hacerse cargo de los retos pedagógicos que presenta el tiempo que nos ha caído en suerte vivir” (Duch, 2004, p. 130). Como vía de solución, él sugiere desarrollar una pedagogía capaz de formular un renovado *humanismo* (Duch, 1997), capaz de desafiar la forma educativa que caracteriza a nuestra época en crisis (Chillón, 2011). Crisis de los conocimientos que velozmente reciben múltiples interpretaciones o “hipótesis revocables y sin ningún tipo de exigencia” (Duch, 2019, p. 42), crisis de la pedagogía que, “sin duda, es muy aguda y con imprevisibles consecuencias para el futuro de nuestra sociedad” (Duch & Mèlich, 2012, p. 202). Por esto, la práctica educativa ha de ser pensada “con el interés de que la tradición cultural no se pierda” (Jiménez, 2023, p. 63). Además, todas las experiencias que poseen significancia para la orientación de la relación y la convivencia, prevalecen únicamente por un sentido óntico que favorece “vivir lo a priori e infieri” (Chillón & Duch, 2012, p. 302).

Con lo señalado, se puede sostener que la mirada de Duch aclara la necesidad de reencontrar en la educación su misión originaria de recibir y formar al ser humano a través de transmisiones auténticas y no simplemente informativas (Moreta, 2019). Puesto que la necesidad más cierta e íntima de la persona es poseer significaciones, para desarrollarse como ser libre y responsable de su trayecto vital al interior de su mundo próximo. El ser humano precisa de una enseñanza basada en conocimientos y prácticas orientadoras para su propio autorreconocimiento dentro del camino que ha

seguido su cultura concreta (Duch, 2004). De ahí la relevancia de la tradición cultural, en su tarea de conservar referencias que unen a las personas en una cosmovisión compartida con la que leen, reinterpretan y traducen su pasado en su presente (Jiménez, 2023). En este sentido, la facultad de la memoria nos recuerda que las personas somos novedades portadoras de historias que han sido tejidas por otros. La memoria nos señala que poseemos una identidad “a partir de aquello que nos viene dado” (Mèlich, 2003, p. 5).

Memoria, pedagogía y cultura

Una de las consecuencias de la actual realidad educativa es la pérdida de las potencialidades de la narración, como medio que exhibe la consciencia cultural e histórica. Duch advierte que no es baladí el curso que la educación adquiere a través de los medios y las formas discursivas que emplea para realizar su tarea de transmisión y formación. Aquí, el autor destaca el serio peligro que impone el monolingüismo del presente, derivado del uso excesivo de las tecnologías en las didácticas y las metodologías educativas. Este lenguaje, piensa Duch, no favorece la comunicación exigida por una vida cultural comunitaria, porque es una forma que desconoce al ser humano desde su estructura *capax symbolorum* (Chillón & Duch, 2012; Duch, 2019). Es decir, como ser simbólico que puede expresarse no sólo con la palabra del “*logos*, es decir, como concepto abstracto, racional, referencial, asensorial y denotativo” (Chillón, 2014, p. 59). Duch explica que debemos asegurarnos de comprender que el ser humano es también *mythos*: imaginación, relato, sensibilidad, emoción (Chillón & Duch, 2012).

El monolingüismo tecnológico no permite formular un sentido existencial captado por la colectividad. De esta manera, el factor comunicacional que interviene en los procesos educativos se desvincula de su función transmisora de contenidos que favorecen la disposición cultural humana, ajustándose la comunicación y la educación a la tecnolatría que es completamente ajena a las formas de la narración cultural. Por esta razón, Duch sostiene que nos encontramos en una crisis pedagógica que repercute hondamente en la capacidad del ser humano para expresar, concretar y *empalabrar* la realidad colmada de “matices y características que no pueden

expresarse con el concurso de una sola forma lingüística” (Duch, 2012a, p. 87).

Debido a la prevalencia del lenguaje informacional que intenta infructuosamente comunicar asuntos profundos de la existencia, la educación y los procesos pedagógicos se han vuelto partes de una cadena de montaje dentro de la globalización. Ciertamente, esto ocurre porque los sistemas educativos no se sostienen en una concepción integral de la inteligencia humana. En otras palabras, la educación posmoderna no aplica la palabra entendida como *logos y mythos*; es decir, como “concepto abstracto e imagen sensorial, como razón y representación, denotación precisa y connotación sensible, referencia analítica y alusión sintética, efectividad y afectividad” (Chillón, 2014, p. 60). En suma, la educación posmoderna no comprende al ser humano en su específica conformación *logomítica*, donde la facultad del recuerdo posee un lugar central.

Consecuentemente, gran parte del problema educativo de nuestra sociedad reside en su distanciamiento de las formas cotidianas de escuchar, hablar y dialogar (Duch, 1997) con saberes y conocimientos previos y pasados. Sin acudir a principios e idearios concebidos colectiva e intergeneracionalmente, la educación posmoderna no sólo reproduce el individualismo, también sobrevalora la aceleración histórica, la discontinuidad y lo transitorio (Jiménez, 2023). Lluís Duch observa que gran parte de la responsabilidad de este fenómeno pertenece al capitalismo que, junto con introducir un agudo individualismo, promueve la creencia en una trascendencia aquí y ahora mediante el hiperconsumo (Chillón & Duch, 2016). Las prácticas sociales pragmáticas y consumistas derivadas del capitalismo marginan a la persona de “su función tradicional de vigorizar el vínculo social y la identidad colectiva” (Duch, 2015, p. 53), al mismo tiempo que socavan la integridad y fiabilidad de las instituciones (Duch, 2019).

Con lo sostenido, se puede decir que la posmodernidad no sólo es incapaz “de plantear una alternativa al macrorrelato triunfante, de cuño neoliberal, tecnocrático y economicista” (Chillón & Duch, 2016, p. 79), también extiende la ausencia de experiencias colectivas. Según nuestro autor,

en la posmodernidad el valor axiológico de un nosotros configurado por la coexistencia entre personas diversas ha sido eclipsado, porque la globalización motiva la homogeneidad cultural y desmotiva el reconocimiento de las acciones que nos vienen del pasado (Duch & Mèlich, 2012). La masificación y la deshistorización (Chillón & Duch, 2012) impulsadas por la posmodernidad, impulsan una forma de vida que no tiene la obligación de acudir a la memoria colectiva y la tradición para obtener un sentido de pertenencia.

Cabe destacar que la tradición aporta sustancialmente a la cultura, porque ella es la memoria colectiva de los momentos a través de los cuales el pasado ingresa en la vida cotidiana. Esta clase de experiencia es la que desarrollaron los pueblos ancestrales tradicionales, que vieron en la memoria un don de los dioses (Duch, 2012b). Por su parte, el ser humano posmoderno se guía por la creencia de que la cultura se produce dentro de una incesante movilidad y cambio de referencias sociales. Sin embargo, Duch considera como aspecto irrevocable que, “la misma sociedad, con sus normas y valores, se inscribe en una determinada memoria vinculante” (Duch, 2012b, p. 178), de otro modo se estropearían por completo las acciones comunicativas y contextualizadoras.

En la perspectiva de Maurice Halbwachs (2004), desde la infancia y gracias al contacto con los adultos, las personas adquieren modos de reencontrar situaciones importantes para ellas. En este lugar, la memoria es esencial para el desarrollo del ser humano como ser social vinculado a otros mediante el intercambio de costumbres, lenguas y tradiciones conocidas. En efecto, memoria y tradición conducen contenidos transgeneracionales que permiten la comprensión cabal de la articulación pasado-presente-futuro (Sola-Morales, 2017, p. 208). Gracias al resguardo de saberes que la memoria proporciona, la persona puede vivir de acuerdo con las condiciones espaciales, temporales e históricas que le son necesarias como sujeto histórico-cultural situado en una realidad con sentido. Más aun, cuando un grupo humano “aborda su pasado, siente que sigue siendo el mismo y toma conciencia de su identidad a través del tiempo” (Halbwachs, 2004, p. 87).

Duch afirma que la sociedad actual dispone de una cantidad innumerable de datos informacionales, vertiginosos y caducos al instante. Esta realidad contribuye al desarrollo de “una civilización del olvido, de la amnesia programada” (Duch, 2020, p. 88); por lo tanto, “constituye uno de los indicadores más importantes de la crisis de lo humano, en general, y, específicamente, de los procesos educativos” (Duch, 2005, p. 138). En este escenario, la pedagogía requiere resignificar el lugar de la memoria en su papel articulador de los saberes que son importantes para las comunidades (Muñoz, 2024). La memoria como premisa de la enseñanza aporta en el reconocimiento de los “sedimentos intersubjetivos que pueden ser reconocidos, recordados y transmitidos por la memoria individual y colectiva como si se tratase de una especie de biografía común de todos los individuos” (Sola-Morales, 2017, p. 209). Porque, como indica Manuel Reyes Mate (2008), no es posible sostener que la razón del ser humano pueda extender sus capacidades cuando es desprovista de la experiencia compartida. Más todavía, una razón sin memoria no permite la valoración del ser humano como ser de relación y comunión.

En esta línea, la cualidad pedagógica de la memoria es que brinda a la razón el desarrollo de una meditación que se esfuerza por superar la sola conceptualización racional, que huye de toda metafísica para desentenderse de la realidad y la tradición (Muñoz, 2024). En suma, la memoria posee importancia, tanto para la persona como para la esfera “transcultural que afecta en profundidad a todos los aspectos significativos de la existencia humana (religión, política, arte, gastronomía, etc.,)” (Duch, 2012b, p. 178).

Reconocer la validez de la memoria significa identificar que su ejercicio no se consigue como función pública de conmemoración; más bien, la memoria ingresa en la formación integral de la persona cuando es asumida como la categoría que conduce a examinar la subjetividad humana y su relación con la historia y el tiempo (Muñoz, 2024). Por lo tanto, la recepción de la memoria como categoría pedagógica en la formación cultural de la persona, es una alternativa para recomponer las consecuencias de la desafectación múltiple que el mundo experimenta.

Revalorar la dimensión de la memoria es una responsabilidad educativa, ética y política que no puede permanecer en la indiferencia. En este sentido. El análisis de Lluís Duch sobre la desestructuración que el presente cultural experimenta y, por consiguiente, en relación a la necesidad de reconsiderar la importancia de los contenidos no exclusivamente conceptuales en la formación del ser humano, no se detiene en un simple diagnóstico del escenario cultural y educativo. Más bien, se extiende hacia una propuesta de comprensión humanista sobre las tramas significativas transportadas a través de la interpretación, la narración, el testimonio, el diálogo y el símbolo como componentes formativos del ser cultural. En este orden, la memoria es imprescindible como facultad que aloja los contenidos que permiten al ser humano ser cultivador de su presente, pasado y futuro.

Conclusiones

Este ensayo ha tenido el propósito de exponer una interpretación de los postulados filosóficos-anropológicos de Lluís Duch, para apreciar la importancia de la formación de la persona como ser cultural. El análisis se ha circunscrito dentro de la observación crítica que Duch mantiene sobre la posmodernidad, a su juicio, desinteresada por el sentido de lo humano que reside en la cultura. Siguiendo al autor, se observa que la precariedad actual de la cultura posmoderna se debe esencialmente a la ausencia de “«memoria cultural»” (Chillón & Duch, 2016, p. 151). Este fenómeno se presenta como una misión de grandes proporciones, pues, mantiene directa relación con los procesos educativos influenciados por el relativismo, la incertidumbre, la provisionalidad, el mercantilismo y el monolingüismo tecnológico que no permite la articulación entre símbolo, narración, diálogo, testimonio y memoria, como aspectos sustanciales de la comunicación. De tal modo, la existencia de la persona en la cultura posmoderna, puede llegar a ser una experiencia asignificativa en términos culturales, debido a la carencia de los relatos biográficamente vinculantes. Lluís Duch esclarece que la especie humana inherentemente dotada para expresar y transmitir saberes y conocimientos, necesita reiteradamente de referencias para acceder a una comprensión significativa de la realidad. Sólo así, es posible la edificación de una biografía situada en un contexto con identidad y sentido. Por lo tanto, es

fundamental experimentar una educación capacitada para su función mediadora de marcos referenciales y comunicadores de la socialización.

Desde la interpretación de Lluís Duch se obtienen fundamentos filosóficos- antropológicos para optar por la memoria como una elección ética y política. En tal sentido, los sistemas educativos podrían asumir las bases esclarecidas por el autor, para cumplir efectivamente como instancias sociales destinadas a transmitir conocimientos no meramente instructivos, sino, formativos y orientadores. Es posible pensar, por ejemplo, en una recuperación de saberes ancestrales que han sido cultivados por las generaciones anteriores y que siempre son parte de la identidad de toda cultura; puesto que, como sostiene José Bengoa (2006), la identidad se relaciona íntimamente con “la “memoria colectiva” que cada grupo posee y que ha podido conservar, o que es capaz de reconstruir en un determinado proceso identitario” (p. 60). Desde aquí, además, es importante destacar la condición *capax symbolorum* del ser humano (Chillón & Duch, 2012), porque su extensión es primordial para establecer una definición de cultura y comunidad.

En síntesis, reconocer la dimensión de la memoria en su carácter antropológico, tal como Duch sostiene, permite desarrollar en el ser humano su propia capacidad proyectiva y sus autodefiniciones, teniendo como base la narración, el diálogo y el testimonio referencial colectivamente compartido. De esta manera, la persona se reconoce no desde el individualismo incrementado por las formas de la posmodernidad, sino, desde una noción de colectividad y pertenencia cultural. En esta línea, la educación, tal como Duch la concibe, posee un papel indudable para la formación de la persona en su necesidad de construir y habitar su mundo con ayuda de las diversas transmisiones que recibirá en su vida.

La ubicación y el crecimiento del ser humano como ser cultural se produce porque él experimenta su mundo como próximo. Esta situación implicará a lo largo de su vida acuerdos y desacuerdos con ciertas nociones; los seres humanos cuestionamos, aceptamos, rechazamos y afirmamos preceptos y normas de toda clase. Pero, en cualquier caso, con mayor o menor

éxito, las personas estamos obligadas a responder personal y colectivamente ante las distintas circunstancias que se nos presentan. Desde aquí, el valor de la memoria es su atributo de “establecer una vinculación creativa con el pasado, a fin de imaginar y configurar el futuro” (Duch, 1997, p. 27).

Referencias

- Bengoa, J. (2006). *La comunidad reclamada. Identidades, utopías y memorias en la sociedad chilena actual*. Catalonia.
- Chillón, A. (2011). *La condición ambigua. Diálogos con Lluís Duch*. Herder.
- Chillón, A. (2014). *La palabra ficticia. Literatura, periodismo y comunicación*. Universitat Autònoma de Barcelona. Servei de Publicacions.
- Chillón, A., & Duch, L. (2012). *Un ser de mediaciones: Antropología de la comunicación* (Vol. 1). Herder.
- Chillón, A., & Duch, L. (2016). *Sociedad mediática y totalismo: Antropología de la comunicación* (Vol. 2). Herder.
- Chillón, A. (2011). La sabiduría antropológica. En pos de un nuevo humanismo. En J.-C. Mèlich, I. Moreta & A. Vega (Coords.), *Empalabrar el mundo. El pensamiento antropológico de Lluís Duch* (pp. 33–50). Fragmenta.
- Choza, J. (2013). *Filosofía de la cultura*. Thémata.
- Duch, L. (1997). *La educación y la crisis de la modernidad*. Paidós.
- Duch, L. (2004). *Estaciones del laberinto*. Herder.
- Duch, L. (2008). La crisis de Dios. *Ars Brevis*, (14), 85–103. <https://raco.cat/index.php/ArsBrevis/article/view/137023>
- Duch, L. (2010). Apología de lo humano: Multidisciplinariedad y Ciencias Humanas. *Ars Brevis*, (16), 11–30. <https://raco.cat/index.php/ArsBrevis/article/view/244253>
- Duch, L. (2012a). *La religión en el siglo XXI*. Siruela.
- Duch, L. (2012b). *Religión y comunicación*. Fragmenta.
- Duch, L. (2015). *Antropología de la ciudad*. Herder.
- Duch, L. (2019). *Vida cotidiana y velocidad*. Herder.
- Duch, L. (2020). *Salida del laberinto*. Fragmenta.
- Duch, L., & Mèlich, J.-C. (2012). *Escenarios de la corporeidad: Antropología de la vida cotidiana* (Vol. 2). Trotta.
- Geertz, C. (2003). *La interpretación de las culturas*. Gedisa.
- Halbwachs, M. (2004). *La memoria colectiva*. Pressas Universitarias de Zaragoza.
- Jiménez, A. (2023). La educación comparada en la historia del tiempo presente. *Pedagogía y Saberes*, (59), 56–68. <https://doi.org/10.17227/pys.num59-17584>

- Lyotard, J.-F. (1987). *La condición posmoderna*. Cátedra.
- Mèlich, J.-C. (2003). Memoria y esperanza. *Comunicação apresentada no IV Encontro de Didática, Cidadania, ensino e aprendizagens da Filosofia*, 171–203. <http://www.apfilosofia.org/documentos/pdf/MelichSant.pdf>
- Mèlich, J.-C. (2008). Antropología narrativa y educación. *Teoría de la educación*, (20), 101–124.
- Moreta, I. (2019). *Conversaciones con Lluís Duch. Religión, comunicación y política*. Fragmenta.
- Muñoz, M. (2024). El lugar de la memoria en *Verdad y método* de Gadamer. *ApareSER. Revista de Filosofía*, 1(2), 62–78. <https://doi.org/10.35494/arfv1i2.18>
- Reyes Mate, M. (2008). *La razón de los vencidos*. Anthropos.
- Rodríguez Gómez, G., Gil Flores, J., & García Jiménez, E. (1996). *Metodología de la investigación cualitativa*. Aljibe.
- Sola-Morales, S. (2017). La importancia de la tradición en los procesos identitarios. *Ars Brevis: Anuario de la Càtedra Ramon Llull Blanquerna*, (23), 207–223. <https://hdl.handle.net/11441/87525>
- Solares, B. (2003). La pedagogía imaginal del mito. En R. Romero Zertuche et al. (Eds.), *Símbolos y arquetipos en el hombre contemporáneo: Formas arquetípicas en la relación maestro/discípulo* (pp. 109–119). Universidad Autónoma Metropolitana. Unidad Xochimilco.
- Solares, B. (2008). Un acercamiento a la antropología simbólica de Lluís Duch. En L. Duch et al. (Eds.), *Lluís Duch, antropología simbólica y corporeidad cotidiana* (pp. 15–40). Centro Regional de Investigaciones Multidisciplinarias-UNAM.

La concepción de la belleza en Tomás de Aquino: Entre lo metafísico y lo divino

The concept of beauty in Thomas Aquinas: Between the metaphysical and the divine

Javier Ignacio Fattah Jeldres

jfattah@ucsc.cl

Universidad Católica de la Santísima Concepción

ORCID ID 0000-0001-8186-2734

Cecilia Soledad Pérez Mora

cperez@ucsc.cl

Universidad Católica de la Santísima Concepción

ORCID ID 0000-0002-9755-9606

Resumen: El pensamiento estético de Tomás de Aquino concibe la belleza no solo como atributo sensible, sino como dimensión metafísica vinculada al ser y a Dios. Aunque la escolástica carece de tratados específicos sobre arte, en Tomás lo bello se integra tanto con la ontología como con la moral, retomando influencias de San Agustín y Dionisio Areopagita. Sus nociones centrales son la proporción, la perfección, la magnitud y la claridad, entendidas tanto en lo material como en lo espiritual. La belleza, en distintos niveles ontológicos, procede de Dios como fuente última, se refleja en el orden del universo y se expresa en el movimiento. En el ámbito moral, un acto bello es aquel en el que sus partes mantienen coherencia, mientras que la fealdad surge de la desproporción o falta de integridad. La *claritas* ocupa un lugar esencial, como luz que proviene desde el alma, el cuerpo y sobre todo desde Cristo, considerado la *perfectissima pulchritudo* al ser imagen fiel del Padre. Así, la estética tomista trasciende lo sensible y lo utilitario, presentando la belleza como propiedad trascendental del ser, inseparable de la verdad, la unidad y la bondad, orientada siempre hacia la divinidad.

Palabras clave: Belleza; proporción; claritas; perfección; Tomás de Aquino.

Abstract: The aesthetic thought of Thomas Aquinas conceives beauty not only as a sensible attribute but also as a metaphysical dimension linked to being and to God. Although scholasticism lacks specific treatises on art, in Aquinas the notion of the beautiful is integrated with both ontology and morality, drawing on the influences of Augustine and Pseudo-Dionysius the Areopagite. His central notions are proportion, perfection, magnitude, and clarity, understood in both the material and the spiritual realms. Beauty, at different ontological levels, proceeds from God as the ultimate source, is reflected in the order of the universe, and is expressed in movement. In the moral sphere, a beautiful act is one in which all its parts maintain coherence, while ugliness arises from disproportion or lack of integrity. *Claritas* holds an essential place, as a light emanating from the soul, the body, and above all from Christ, considered the *perfectissima pulchritudo* as the faithful image of the Father. Thus, Thomistic aesthetics transcends the sensible and the utilitarian, presenting beauty as a transcendental property of being, inseparable from truth, unity, and goodness, and always oriented toward the divine.

Keywords: Beauty; proportion; claritas; perfection; Thomas Aquinas.

Introducción

Este artículo tiene como objetivo principal examinar las nociones estéticas fundamentales en el pensamiento de Tomás de Aquino, haciendo énfasis en su concepción metafísica de la belleza como participación en el ser divino.

La metodología adoptada es de carácter hermenéutico y comparativo. Se realiza una exégesis filosófica de los principales textos de Tomás de Aquino (como la *Summa Theologiae*, *Scriptum super Sententiis*, *Summa Contra Gentiles* y Comentarios a Dionisio), atendiendo a su contexto escolástico.

Las nociones estéticas han estado presentes a lo largo de los diversos periodos de la historia. No obstante, existen épocas en las que estas se han manifestado con mayor claridad y profundidad. En el caso particular de la escolástica, no es posible afirmar que existan tratados específicos sobre el arte, al menos no de manera directa. Sin embargo, la idea de lo estético sí está presente, como se demostrará a lo largo del desarrollo de este trabajo. Aunque dentro de la escolástica se pueden identificar diversos autores de importancia considerable, en este escrito nos centraremos en el pensamiento tomista, dado que el Doctor Angélico retoma nociones estéticas de autores previos como San Agustín, Alberto Magno y Dionisio de Areopagita.

En las concepciones estéticas del dominico, es evidente que la belleza no se limita al ámbito material, sino que también adquiere un marcado carácter metafísico. Por ello, el autor presta especial atención a la belleza del alma y a la proporción divina. En este sentido, es posible identificar cuatro conceptos fundamentales en su pensamiento: la proporción (*proportio*), la magnitud (*magnitudo*), la perfección (*integritas sive perfectio*) y la claridad (*claritas*). Estas cuatro características se encuentran presentes tanto en el ámbito material como en el espiritual, razón por la cual Tomás de Aquino pone énfasis en la creación divina e, incluso, en la belleza intrínseca de Dios y del Hijo.

Debido al carácter eminentemente metafísico de las nociones estéticas en Aquino, este establece diferentes niveles ontológicos de la belleza. En el primer nivel se encuentra la belleza divina, de la cual emanan todos los demás tipos de belleza. En un segundo nivel ontológico se ubica el orden, que, según el autor, se manifiesta en el universo. Por último, como tercer nivel ontológico de la belleza, se encuentran el movimiento y el descanso.

Además, la belleza desempeña un rol moral en el pensamiento tomista. De manera análoga a lo que sucede con los cuerpos, Santo Tomás considera que un acto moral puede ser calificado como bello en la medida en que cada una de sus partes contribuye a la plenitud del conjunto. En otras palabras, un acto moral sería bello cuando todas sus partes poseen coherencia y excelencia.

A pesar de la relevancia de conceptos como la proporción, la magnitud y la perfección en la estética tomista, consideramos que el concepto de *claritas* es uno de los más significativos. Este concepto se entiende como una luz que no solo se manifiesta en Dios, sino también en el Hijo. Incluso, Santo Tomás llega a afirmar que en el alma humana existe *claritas*, la cual trasciende hacia el cuerpo, iluminándolo.

Por último, debemos señalar que la hipótesis central de este trabajo es que la estética tomista, articulada en torno a *integritas*, *proportio* y *claritas*, constituye una estructura metafísica que permite comprender la belleza como una propiedad del ser y no solo como fenómeno sensible, y que además esta estructura puede seguir siendo válida para el debate contemporáneo.

El contexto escolástico y la noción de arte

Antes de cualquier aproximación al pensamiento estético de Santo Tomás de Aquino, es imperativo considerar una cuestión fundamental: El doctor Angélico pertenece al siglo XIII, es decir, a la época escolástica. En consecuencia, dentro del contexto histórico en el que se desarrolla su obra, no es posible encontrar un tratado sobre el arte entendido de manera autónoma. Más bien, el arte es abordado de manera general o como una dimensión subordinada a otros temas.

Esto, sin embargo, no implica en absoluto que los escolásticos carecieran de nociones o reflexiones sobre el arte. Por el contrario, autores como Jacques Maritain han subrayado este punto en sus análisis sobre la relación entre el arte y la filosofía escolástica:

Los escolásticos no han escrito tratado alguno especial intitulado Filosofía del Arte. Sin embargo, se encuentra en ellos un a teoría del arte muy profunda, pero hemos de buscarla en las disertaciones austeras sobre algún problema de lógica de si la Lógica es un arte liberal", o de teología moral —"¿cómo la virtud de la Prudencia, virtud a la vez intelectual y moral, se distingue del Arte, que es una virtud intelectual?" (Maritain, 1972, p. 7)

Como hemos señalado, a pesar de la ausencia de un tratado específico sobre el arte en el contexto de la escolástica, es posible encontrar una comprensión de este concepto. El arte se considera una virtud intelectual. Dado que las virtudes son hábitos operativos que conducen al bien, podemos concluir que los escolásticos otorgaban al arte una connotación positiva. En relación con esto, Santo Tomás de Aquino afirma: “La virtud, según se ha dicho anteriormente (q.55 a.3), es un hábito por el cual se obra bien” (Aquino, 1989, p. 429).

Es importante destacar que, en el pensamiento escolástico, el concepto de arte no se restringe exclusivamente a la pintura o la escultura, sino que se aborda desde una perspectiva mucho más amplia. Los escolásticos consideraban a los artesanos y carpinteros como artistas en su propio ámbito. En otras palabras, no establecían una distinción marcada, como la que existe en la actualidad, entre lo que conocemos como bellas artes y artesanía.

En este contexto, es válido afirmar que los escolásticos no percibían una diferencia sustancial entre la utilidad y la belleza:

A decir verdad, la división de las artes en artes de lo bello (bellas artes) y artes de lo útil, por más importante que sea, no es lo que los Lógicos llaman una división 'esencia'; se funda en el fin perseguido, y un mismo arte puede muy bien perseguir a la vez la utilidad y la belleza. Tal es por excelencia el caso de la arquitectura. (Maritain, 1972, p. 159)

Ahora, dentro del marco de la discusión estética contemporánea, podemos destacar que autores como, Jan Aertsen, o Hans Urs von Balthasar han replanteado la tríada tomista de *integritas*, *proportio* y *claritas* como un modo de comprender la experiencia estética más allá del subjetivismo moderno.

Por ejemplo, Aertsen considera que la *integritas* puede ser entendida hoy como la plenitud estructural del objeto; la *proportio* como relación inteligible que hace posible la forma, y la *claritas* como condición de manifestación, acercándose de este modo a nociones fenomenológicas. Esto permite evidenciar que la belleza en Tomás no es meramente simetría o luminosidad, sino una forma de inteligibilidad ontológica.

Las fuentes de la estética tomista

Una vez que hemos establecido algunas consideraciones fundamentales sobre el arte en el contexto escolástico, podemos adentrarnos en el pensamiento del Dominico. En este sentido, uno de los primeros aspectos a considerar son las influencias estéticas que moldearon su enfoque, especialmente aquellas provenientes del mundo medieval y, por ende, del cristianismo. Esto no excluye en absoluto la influencia del pensamiento griego en su obra; sin embargo, con el objetivo de centrarnos en lo específico de la época medieval, nos enfocaremos en estas últimas.

Uno de los autores que influyó notablemente en Santo Tomás fue Agustín de Hipona. De hecho, al analizar los textos del Aquinate, encontramos que este reconoce de manera explícita que algunas de las ideas que expresó sobre cuestiones estéticas tienen sus raíces en San Agustín, algo que se evidencia en diversos pasajes de la Suma de Teología “pues Agustín, en *Enchirid.*, dice que la admirable belleza existe a partir de la totalidad” (Aquino, 2001, p. 472).

Las referencias a las obras de San Agustín en cuestiones relacionadas con la estética son una constante, incluso en temas relacionados con el cuerpo, donde Santo Tomás de Aquino recurre a Agustín para explicar la belleza de los cuerpos. Por ejemplo, cuando afirma que:

Pero los cuerpos gloriosos estarán dotados de color, porque, como dice Agustín en *La Ciudad de Dios*, la belleza del cuerpo es la proporción adecuada de las partes con cierta suavidad del color. Ahora bien, la belleza no podrá faltar en los cuerpos glorificados. Luego, los cuerpos gloriosos no serán luminosos. (Aquino, 1895, p. 370)

Otra influencia significativa en la estética del Doctor Angélico es la ejercida por Dionisio de Areopagita, a quien Aquino cita de manera constante e incluso realiza comentarios sobre su obra *Los Nombres Divinos*. En este contexto, es importante señalar que la influencia de Dionisio no parece limitarse al ámbito estético, aunque es en este campo donde se puede afirmar que ejerce una influencia aún mayor que en el resto de la obra del Aquinate. En esta línea, señala Artesen: “The most important authority for Thomas's concept of beauty is pseudo Dionysius the Areopagite” (Artesen, 1991, p. 78).

Los elementos constitutivos de la belleza

Después de mencionar algunas de las fuentes relacionadas con la estética tomista, podemos adentrarnos en los conceptos propios del autor. En este sentido, debemos comenzar señalando que, según Tomás de Aquino, existen condiciones que podemos considerar esenciales para la belleza. Entre ellas se encuentra la noción de proporción, la cual no es exclusiva de Santo Tomás, ya que también está presente en las obras de Dionisio y San Alberto Magno.

Para Aquino, la proporción no se limita a un solo ámbito, sino que se manifiesta de maneras diferentes según el tipo de ente. Un ejemplo de ello es la consideración de los seres humanos, quienes se encuadran en lo que podemos denominar un ente particular corpóreo. En este caso, la idea de belleza está estrechamente ligada a una cuestión física, en el sentido de que se asocia la belleza a la proporción entre las partes del cuerpo:

Tal como podemos deducir de las palabras de Dionisio en *IV De Div. Nom.*, concurren en la noción de bello o decoroso el brillo y la proporción debida. En efecto, según él, Dios es llamado bello como causa de la armonía y del brillo del universo. Por eso la belleza del cuerpo consiste en que el hombre tenga los miembros corporales bien proporcionados, con un cierto esplendor del color conveniente. (Aquino, 1994, p. 423)

Hemos identificado la belleza con la existencia de la proporción. Siguiendo la doctrina tomista, es posible relacionar la falta de proporción con

lo que podríamos considerar como lo "feo". Sin embargo, esta idea de proporción se aplica únicamente a los entes de carácter físico. Es evidente que el alma carece de materialidad y, por tanto, no puede albergar ni armonía ni complexión: “Todo esto demuestra que el alma ni es armonía ni complexión” (Aquino, 1952, p. 558).

Queda claro que la belleza y, por lo tanto, la proporción, no se limitan necesariamente a la materia. Esto nos permite afirmar que la proporción puede manifestarse en entes particulares de carácter incorpóreo. Uno de los ejemplos más claros es el caso de los ángeles. Aquino sostiene:

Además, en los ángeles está la máxima belleza después de Dios. Pero la razón de la belleza consiste en la armonía ordenada de diferentes grados. Por lo tanto, parece que en los ángeles también se puede encontrar una gradación de este tipo dentro de un mismo orden. (Aquino, 1856, p. 464)

Es importante destacar que la belleza no se limita estrictamente a los entes, sino que también puede encontrarse en las cualidades de los entes. Es posible identificar la belleza en el ámbito moral o incluso hablar de una belleza de carácter espiritual, que, sin duda, se sitúa por encima de una belleza física o exterior.

Un tercer sentido en el que las nociones estéticas se manifiestan, a juicio de Tomás de Aquino, es la proporción del universo, la cual posee un notorio carácter metafísico. En este sentido, la idea de proporción en el universo se encuentra ligada a la idea de Dios, en tanto que aquello que proviene de la creación divina manifiesta la belleza “Por eso, también en el orden de lo producido por las criaturas se manifiesta la belleza de la sabiduría divina” (Aquino, 2001, p. 364).

El dominico tiende a identificar la relación de proporción entre Dios y su creación con la *habitus ordinis*. Sin embargo, si la proporción del universo se identifica con Dios, se podría pensar que solo aquello que es bueno es parte

de lo bello. No obstante, Aquino reconoce en la Suma contra Gentiles que lo malo contribuye a la posibilidad de admirar la belleza del universo:

Pero si se suprimiera el mal de algunas partes del universo, se perdería mucho de su perfección, porque su belleza se nos muestra por la ordenada conjunción de males y de bienes, ya que los males provienen de los defectos de los bienes y, esto no obstante, su resultado es la aparición de ciertos bienes por la providencia del gobernador, tal como la interposición de silencios hace placentera la melodía. Por lo tanto, la divina providencia no debió excluir el mal de las cosas. (Aquino, 1952, p. 292)

En Santo Tomás, la posibilidad de abordar ciertos aspectos de la metafísica en relación con el universo y la belleza resulta evidente por dos motivos fundamentales. El primero de estos motivos se vincula directamente con Dionisio, ya que, en los comentarios que el Aquinate realiza sobre la obra de Dionisio, reconoce que la consonancia puede manifestarse en un sentido doble. Desde esta perspectiva, y como primer argumento, se puede sostener que la consonancia está directamente relacionada con Dios, considerándolo como su causa directa, quien, además, orienta todas las cosas hacia sí mismo como fin. Ahora bien, si consideramos la consonancia en un segundo sentido, podemos comprenderla en relación con el orden que las cosas mantienen respecto a otras. De este modo lo expresa Santo Tomás:

Rursus exponit aliud membrum, scilicet quod Deus sit causa consonantiae in rebus; est autem duplex consonantia in rebus: prima quidem, secundum ordinem creaturarum ad Deum et hanc tangit cum dicit quod Deus est causa consonantiae, sicut vocans omnia ad seipsum, inquantum convertit omnia ad seipsum sicut ad finem, ut supra dictum est et propter hoc pulchritudo in Graeco callos dicitur quod est a vocando sumptum; secunda autem consonantia est in rebus, secundum ordinationem earum ad invicem; et hoc tangit cum

subdit, quod congregat omnia in omnibus, ad idem.¹ (Aquino, 1950, p. 338)

Un segundo elemento que nos permite abordar la noción de belleza en un sentido metafísico está relacionado con la existencia de lo que podemos denominar niveles ontológicos en relación con la consonancia que exhibe la belleza. En este sentido, el primer nivel ontológico es la belleza divina, entendida como aquella de la cual emana toda consonancia. En este contexto, se postula que la belleza divina actúa como la fuente primordial de armonía.

El segundo nivel ontológico guarda relación con el orden, manifestándose en la existencia de cierta armonía en el universo. Parece existir un orden intrínseco, donde los entes tienden hacia un fin de manera ordenada.

En tercer lugar, se encuentra el movimiento y el descanso. El descanso, al igual que el movimiento, presenta un orden. Es evidente que los seres se mueven siguiendo ciertos patrones, aunque ello no implica necesariamente una causa externa del movimiento, al menos no en los seres racionales. En este caso, los seres racionales poseen un movimiento interno, siendo su propia causa, pero están orientados hacia un fin, lo que implica cierto orden en su movimiento.

Finalmente, en estrecha relación con los demás tipos de proporción, se halla la proporción divina. Esta puede entenderse como la proporción fundamental, especialmente si aceptamos la premisa de que toda consonancia proviene de la belleza divina, como afirma Santo Tomás de Aquino: “Similiter etiam dictum est quod de ratione pulchritudinis est consonantia, unde omnia,

¹ Expone de nuevo otro aspecto, a saber, que Dios es la causa de la consonancia (armonía) en las cosas; hay, sin embargo, una doble consonancia en las cosas. La primera, ciertamente, según el orden de las criaturas hacia Dios; y alude a ella cuando dice que Dios es causa de la consonancia, en cuanto llama a todas las cosas hacia sí mismo, en la medida en que las convierte a sí mismo como a su fin, como se dijo más arriba. Y por esto, la belleza, en griego, se dice kálos, lo cual procede de “llamar” (kálein). La segunda consonancia, en cambio, está en las cosas según la ordenación de ellas entre sí; y a esto alude cuando añade que Él reúne todas las cosas en todas, hacia lo mismo (Traducción propia).

quae, qualitercumque ad consonantiam pertinent, ex divina pulchritudine procedunt²” (Aquino, 1950, p. 309).

Como ya hemos señalado, la belleza consiste en la proporción. En el caso de Dios, se reconoce que, al menos según la obra de Santo Tomás, la belleza presente en lo divino está conectada con la Trinidad. Esto se evidencia en los comentarios a las Sentencias: “Perfectissima pulchritudo intelligitur filius” (Aquino, 1856, p. 309) donde se presenta al Hijo como semejante a la imagen divina. En este sentido, es posible afirmar que, a medida que una representación se acerca a lo que pretende representar, adquiere un mayor grado de belleza. Jesús, en este contexto, encarna la belleza de manera clara puesto entre todos los seres resulta ser el más cercano a la imagen de Dios. Cabe destacar que la belleza divina puede entenderse en un sentido doble: como brillo y proporción, según lo expresa Santo Tomás: “Pulchritudo consistit in duobus, scilicet in splendore, et partium proportione³” (Aquino, 1856, p. 309).

A este doble sentido, y en concordancia con lo señalado, es posible agregar un tercer elemento: la verdad. En la medida en que una imagen es fiel a lo que pretende representar, se puede afirmar que posee un mayor o menor grado de belleza, en virtud de ser una representación más verdadera.

Como se ha indicado previamente en este trabajo, diversas cualidades se asocian con la noción de lo bello, como es el caso de la proporción, la cual ya ha sido explicada. Otra cualidad relevante es el tamaño, la *magnitudo*. En este punto, es importante destacar que Santo Tomás identifica la magnitud con la *integritas sive perfectio*, al menos así parece indicarlo la comparación entre las Sentencias de Pedro Lombardo y la primera parte de la Suma Teológica. En el caso de las Sentencias, se observa que entre las tres condiciones de la belleza se incluye la *magnitudo*:

² De manera similar, también se ha dicho que la razón de la belleza es la consonancia, por lo que todas las cosas que de cualquier modo pertenecen a la consonancia, proceden de la belleza divina (Traducción propia).

³ La belleza consiste en dos cosas, a saber, en el esplendor y la proporción de las partes (Traducción propia).

Ad rationem autem pulchritudinis duo concurrunt, secundum Dionysium, scilicet consonantia et claritas. Dicit enim, quod Deus est causa omnis pulchritudinis inquantum est causa consonantiae et claritatis, sicut dicimus homines pulchros qui habent membra proportionata et splendentem colorem. His duobus addit tertium philosophus ubi dicit, quod pulchritudo non est nisi in magno corpore; unde parvi homines possunt dici commensurati et formosi, sed non pulchra.⁴ (Aquino, 1856, p. 724)

De acuerdo con lo expresado en los *Comentarios a las Sentencias*, observamos que el término "tamaño" aún se utiliza como una de las condiciones de la belleza. Sin embargo, en textos posteriores, como la Suma Teológica, dicho término no aparece, sino que se emplea la expresión "integridad" o "perfección". De esta manera, podemos concluir que existe una cierta identificación entre la *magnitudo* y la *integritas sive perfectio*:

La especie o la belleza tienen semejanza con lo propio del Hijo, pues para la belleza se requiere lo siguiente: Primero, integridad o perfección, pues lo inacabado, por ser inacabado, es feo. También se requiere la debida proporción o armonía. Por último, se precisa la claridad, de ahí que lo que tiene nitidez de color sea llamado bello. (Aquino, 2001, p. 389)

Aclarada la continuidad que existe entre *magnitudo* e *integritas sive perfectio*, podemos explicar que, al igual que sucede con la proporción, la integridad se manifiesta de diversos modos, siendo un primer caso la integridad en el ente de carácter particular.

⁴ Según Dionisio, en la noción de belleza concurren dos aspectos, a saber, la consonancia y la claridad. Se dice, pues, que Dios es la causa de toda belleza en cuanto es causa de la consonancia y la claridad, tal como decimos que los hombres son hermosos cuando tienen miembros proporcionados y un color resplandeciente. A estos dos, el Filósofo añade un tercero, cuando dice que la belleza no se encuentra más que en un cuerpo grande. Es por tanto que los hombres pequeños pueden ser llamados proporcionados y bien formados, pero no bellos (Traducción propia).

La integridad en los entes corpóreos guarda una relación directa con una cuestión física. En este sentido, si a un individuo le faltara una parte del cuerpo, no podríamos considerarlo bello. Aunque es cierto que, en el caso de lo corpóreo, puede resultar más fácil identificar la integridad, nuestro autor es consciente de la existencia de una belleza de carácter espiritual, la cual resulta análoga a la belleza corporal, como se expresa en *Contra Impugnantes*:

Est enim duplex pulchritudo: una spiritualis quae consistit in ordinatione debita animae et affluentia bonorum spiritualium, unde omne quod procedit ex defectu spiritualis boni vel quod interiorum inordinationem ostendit turpitudinem habet.⁵ (Aquino, 1902, p. 176)

La dimensión moral de la belleza

Es evidente que la belleza no se limita únicamente a lo físico. Por lo tanto, podemos afirmar que también puede existir belleza en el ámbito moral. La noción de belleza en el acto moral resulta particularmente intrigante. Al igual que ocurre con otros tipos de belleza, se requiere que cada uno de los elementos que conforman el acto moral sea bello para que podamos considerarlo en su totalidad como bello. De este modo, podemos observar esto en las *Cuestiones Disputadas* acerca de lo malo:

Mas como dice Dionisio "el bien es producto de una causa única e íntegra, en cambio el mal resulta de cualquier defecto particular". Y por esto, cuando sea malo, ya sea el acto o la ordenación del acto a su fin, entonces se considerará el todo como malo. En cambio el todo no se considera bueno a no ser que ambas cosas sean buenas: por ejemplo, no se considera bello a un hombre a no ser que todos sus miembros sean bellos; mas se le considera feo con que uno solo de sus miembros sea feo. Y de aquí que el acto malo no puede llegar a ser bueno, pues debido a que el acto es malo, no puede ser bueno íntegramente; en cambio el acto bueno puede llegar a ser malo porque

⁵ Existe, pues, una doble belleza: una espiritual que consiste en la debida ordenación del alma y la abundancia de bienes del espíritu, de manera que todo lo que procede del defecto de un bien espiritual o lo que ostenta desorden interior, tiene fealdad (Traducción propia).

para ser malo no se requiere que sea malo íntegramente, sino que basta que sea malo en algún aspecto particular. (Aquino, 2015, p. 99)

Si bien es cierto que basta con que falte una de las partes para que podamos considerar un acto moral o un cuerpo como feo, no debemos dejar de reconocer que, según la parte que falte, se puede hablar de un mayor o menor grado de fealdad. En otras palabras, en la noción de integridad existe cierta graduación. Siguiendo la analogía establecida entre las partes del cuerpo y el acto moral, utilizaremos el ejemplo del pecado para evidenciar que no toda falta moral, como es el caso del pecado, supone el mismo grado de fealdad. Así, podemos considerar que el pecado de omisión y el pecado mortal no se encuentran en el mismo nivel:

Así, pues, no toda deformidad del cuerpo es igual, sino que alguna es mayor que otra, de modo que las cosas que pertenecen al decoro o a algo más importante son privadas en mayor cantidad, de modo que no toda deformidad o desorden de un acto es igual, sino que alguno es mayor que otro. (Aquino, 2015, p. 129)

Conviene notar que, al igual que sucede con el cuerpo, la noción de naturaleza es fundamental para el acto moral y, por consecuencia, para su belleza. En este sentido, el acto moral realizado por un individuo se considerará como positivo en la medida en que esté acorde con su naturaleza.

En otro ámbito en el que podemos identificar la integridad es en el universo; es decir, el universo exhibe un orden. Sin embargo, este orden no es único, sino que se manifiesta de manera dual. El primer tipo de orden se refiere a la relación entre las partes del universo, mientras que el segundo orden está vinculado a la orientación del universo hacia Dios. No obstante, y a pesar de las diferencias que puedan presentarse, es imperativo que exista integridad, ya que se refiere a todo lo que existe. Al referirse a todo lo que existe, es evidente que también se incluyen elementos que pueden considerarse como negativos. De hecho, nuestro autor reconoce explícitamente que la belleza del universo está relacionada con la conjunción de bienes y males, como se ha señalado anteriormente.

Una tercera característica que resulta esencial para la noción de belleza en Santo Tomás es la idea de perfección, la cual no es exclusiva de Tomás, sino que es posible encontrarla en autores anteriores. No obstante, aceptando la influencia que puedan haber tenido autores precedentes en las concepciones de Santo Tomás, debemos considerar que, en él, la perfección que puede tener la belleza, a nuestro juicio, se manifiesta en tres vertientes esenciales. En el primer caso, encontramos aquella perfección que se expresa en el plano ontológico, en el ser mismo, siendo quizás uno de los casos más destacados el de los ángeles, ya que estos se inclinan al pecado como consecuencia de la belleza: “et sic primum peccatum angeli superbia fuit: quod patet tum ex desiderio, quia eminentiam dignitatis appetiit: tum etiam ex motivo, quia ex consideratione propriae pulchritudinis in peccatum ruit⁶” (Aquino, 1856, p. 282).

En la segunda vertiente, identificamos el conocimiento, en el sentido de que, en la medida en que un determinado objeto es bello, podemos asumir que el conocimiento que lo ha producido es también más perfecto. Ahora bien, se puede inferir que, en el caso de la belleza en el ser humano, al ser creado por Dios, debería ser completamente bello. Sin embargo, debido al pecado original, ha perdido tanto belleza como conocimiento:

Se dice, en el mismo libro, que la disminución de belleza en la naturaleza humana se imputa como pecado. Pero la máxima belleza de la naturaleza humana consiste en el resplandor del conocimiento. Luego parece que el pecado original, que se imputa a la naturaleza humana, es una disminución de conocimiento, es decir, ignorancia. (Aquino, 2015, p. 241)

La última vertiente de la perfección que resulta conveniente identificar es la perfección en Dios, la cual se contempla de tres modos diferentes. En primer lugar, se relaciona con la potencia; en segundo lugar, con la belleza, razón por la cual Dios puede ser considerado como luz, pero

⁶ Y así el primer pecado del ángel fue la soberbia: lo que queda claro tanto por el deseo, ya que buscó la eminencia de la dignidad, como también por el motivo, ya que, a partir de la consideración de su propia belleza, cayó en el pecado (Traducción propia).

en un sentido absoluto, dado que en la luz divina no existe composición. Como tercera y última manifestación de su perfección, encontramos que su naturaleza es esencialmente espiritual. En relación con esta característica, Santo Tomás señala:

Quantum ad bene esse consideratur divina perfectio quantum ad tria. Primo quantum ad potestatem, et secundum hoc ipse Deus dicitur virtus et non tantum habens virtutem. Unde non potest esse quod componatur ex aliquibus, quia in componentibus non esset virtus illa, et ita componeretur ex infirmis, idest carentibus virtute quae est totius. Virtus autem non habet aliquid admixtum, sicut nec esse; sed habens virtutem potest habere infirmitatem admixtam. Secundo quantum ad pulcritudinem, et sic Deus dicitur lux. Si autem ista lux sequeretur compositionem aliquam, componentia non essent lucentia; sicut videmus quia ex congregatione diaphani efficitur corpus lucidum, ut chrySTALLUS, cum tamen partes quas dispersit, non haberent prius luciditatem. Nihil autem obscurum in Deo potest esse, qui est lux, sicut nec non esse in esse. Tertio quantum ad naturae subtilitatem, dicitur enim spiritus essentialiter. Unde formalitas sua, sive quidditas, non est ex aliquibus disparibus, id est dissimiliter se habentibus ad formalitatem, sicut humanitas ex anima et corpore, per quam homo formaliter est.⁷ (Aquino, 1856, pp. 235-236)

⁷ En cuanto al bien ser, se considera la perfección divina en tres aspectos. Primero, en cuanto al poder, y en este sentido Dios es llamado virtud y no sólo poseedor de virtud. Por tanto, no puede ser que Dios esté compuesto de dos partes, ya que en los componentes no habría esa virtud que es, y así se compondría de elementos débiles, vale decir, carentes de la virtud que es propia del todo. La virtud, sin embargo, no tiene nada mezclado en ella, como tampoco el ser. Pero el que posee virtud puede tener debilidad mezclada. Segundo, en cuanto a la belleza, Dios es llamado “Luz”. Si esta Luz estuviere consujeta a alguna composición, los componentes no serían luminosos. Como vemos que de la unión de lo diáfano se forma en un cuerpo luminoso, como el cristal, aunque las partes dispersas no tienen luminosidad previa. Sin embargo, nada obscuro puede haber en Dios, que es luz, como tampoco puede haber “no-ser” en el ser. Tercero, en cuanto a la sutileza de la naturaleza, Dios es esencialmente llamado espíritu. Por ello, su formalidad, su “quidditas”, no está hecha de partes dispares, vale decir, de elementos que se relacionan desigualmente con la formalidad,

Finalmente, podemos mencionar el concepto de *claritas*, el cual resulta esencial en las concepciones estéticas de Santo Tomás. Al igual que ocurre con los conceptos tratados previamente, la *claritas* posee diversos matices. No obstante, para ofrecer una definición preliminar, podemos afirmar que la *claritas* representa una especie de luz. En este sentido, y como lo reconoce Tomás, existe una relación entre lo que podemos denominar la luz del cuerpo y la luz del alma, en el sentido de que esta última tiene la capacidad de manifestarse a través de la primera, como ocurrió en el caso de Cristo.

De esta afirmación, podemos inferir que la *claritas* puede ser dividida en dos tipos: corpórea e incorpórea. En el primer caso, es decir, en el caso de la *claritas* corpórea, podemos afirmar que lo que otorga un determinado grado de belleza a un ente es la claridad, tal como se expresa en los comentarios al libro de Job: “Subdit autem de his quae habent pretiositatem ex pulchritudine dicens non adaequabitur ei aurum, quod habet pulchritudinem ex splendore, vel vitrum, quod habet pulchritudinem ex perspicuitate quamvis non sit praececellens pretii aestimatione”⁸ (Aquino, 1965, p. 28).

Si bien la *claritas* es esencial para la belleza en los entes corpóreos, no se encuentra aislada, sino que está vinculada a la idea de color, en el sentido de que un ente es bello en la medida en que presenta una adecuada distribución de colores. La existencia de la *claritas* en los entes corpóreos es indispensable para la belleza, hasta el punto de que nuestro autor llega a considerar que los muertos carecen de belleza debido al proceso de decoloración que experimentan. Esta idea está relacionada con las "cualidades" que el alma transfiere al cuerpo, de modo que podemos inferir que la ausencia del alma en el cuerpo lo hace menos bello:

Tertio dat pulchritudinem et claritatem; infirmi enim et mortui propter debilitatem operationis animae in corpus, efficiuntur

como la humanidad, la que está formada por alma y cuerpo, a través de los cuales el hombre es formalmente tal (Traducción propia).

⁸ Y añade sobre aquellas cosas que tienen preciosidad debido a su belleza, diciendo que no se le igualará el oro, que tiene belleza por su resplandor, ni el vidrio, que tiene belleza por su transparencia, aunque no sean superiores en cuanto la a estimación de su valor (Traducción propia).

discolorati, et quando erit in summa perfectione, faciet corpus clarum et fulgidum. Quarto dat motum, et tanto facilius, quanto virtus animae fuerit fortior supra corpus. Et ideo quando erit in ultimo suae perfectionis, dabit corpori agilitatem.⁹ (Aquino, 1953, p. 110)

En virtud de lo señalado, la *claritas* no solo está presente en lo corpóreo, sino también en aquello que es incorpóreo, como en el caso del alma, la cual ilumina al cuerpo. A esto, y considerando lo ya mencionado acerca de la relación de la belleza con la moral, debemos hacer una breve mención de la *claritas* presente en la belleza moral, la cual es entendida por nuestros autores en un sentido doble. Es decir, existe un fulgor interior y otro exterior; en ambos casos, se establece una relación con el pecado y, por consiguiente, con la moral. Al respecto, el Aquinate señala:

así también en el alma hay un doble esplendor: el uno habitual, como un fulgor interior; y el otro actual, como un fulgor exterior. El pecado venial impide el fulgor actual, mas no el habitual: porque no excluye ni disminuye el hábito de la caridad y de las otras virtudes, como se verá más abajo (2-2 q.24 a.10; q.133 a.1 ad 2), sino que sólo impide su acto. (Aquino, 1989, p. 686)

Por último, es necesario señalar la existencia de la *claritas* divina, la cual está relacionada con la belleza en Dios. En este contexto, es importante establecer una distinción entre la belleza de Dios y la del Hijo. En el primer caso, es decir, en la belleza de Dios, esta es una parte esencial de Él y no un añadido:

Deus enim non habet circumdatum decorem quasi superadditum eius essentiae sed ipsa essentia eius est decor, per quem intelligitur ipsa claritas sive veritas, et puritas sive simplicitas, et perfectio essentiae

⁹ Tercero, da belleza y claridad; porque los débiles y muertos, debido a la debilidad de (la) acción del alma en el cuerpo se tornan descoloridos, y cuando esté en su máxima perfección, hará el cuerpo claro y resplandeciente. Cuarto, da movimiento, y tanto más fácilmente cuanto más fuerte sea la virtud del alma sobre el cuerpo. Y por eso, cuando esté en la última de sus perfecciones, dará al cuerpo agilidad (Traducción propia).

eius; sed homo non potest habere decorem nisi circumdatum, quasi participando ipsum a Deo ut superadditum suae essentiae.¹⁰ (Aquino, 1965, p. 40)

Podemos concluir que, al ser Dios la belleza en sentido absoluto parece evidente que toda belleza presente en los entes emana de Él de alguna manera. Los seres adquieren un mayor grado de belleza en la medida en que participan de la esencia divina. Habiendo señalado algunos aspectos de la belleza en Dios, podemos abordar la belleza del Hijo, la cual procede de la belleza del Padre. La belleza del Hijo no se limita a la armonía, sino que está intrínsecamente relacionada con el Verbo. Cristo es bello porque es el Verbo, es decir, está en acto; por lo tanto, no carece de ninguna cualidad.

Conclusiones

La estética en Tomás de Aquino constituye una de las expresiones significativas del pensamiento medieval, en tanto logra articular lo bello con la ontología, la moral y la teología. A pesar de que su obra no contempla un tratado específico sobre estética o arte en sentido moderno, su sistematización filosófica permite deducir una concepción rigurosa de lo bello, tanto en el plano material como en el espiritual. En el autor la belleza no se presenta como una categoría accesoria, sino como un reflejo del orden mismo del ser, una propiedad que, junto con la unidad, la bondad y la verdad, constituye una de las trascendentales del ente.

A lo largo del presente trabajo hemos visto que Tomás de Aquino elabora una concepción jerárquica y analógica de la belleza, en la cual lo sensible y lo inteligible participan de un mismo fundamento último: la perfección divina. Desde esta perspectiva, se puede afirmar que toda belleza es participación, ya sea mediante la proporción armónica de las partes, la claridad luminosa del ser o la integridad de aquello que está conforme con su

¹⁰ Porque Dios no tiene una belleza circundante con algo añadido a su esencia, sino que su misma esencia es belleza, por medio de la cual se entiende su claridad o verdad y su pureza o simplicidad, y la perfección de su esencia. Pero el hombre no puede tener belleza más que la circundante, como participando de ella de Dios como algo añadido a su esencia (Traducción propia).

naturaleza. Estos tres elementos *proportio*, *claritas* e *integritas sive perfectio* son retomados de la tradición agustiniana, del pensamiento neoplatónico de Dionisio Areopagita, y de su propia reelaboración aristotélica. Así, Tomás actualiza estos conceptos dentro del horizonte teológico del siglo XIII, en el que lo creado no es autónomo sino imagen del Creador.

Uno de los grandes logros del pensamiento tomista consiste en extender el análisis de la belleza más allá de lo estético en sentido estricto, hacia lo moral y lo metafísico. Esto permite sostener que la belleza no se limita a las obras de arte o a los cuerpos proporcionales, sino que se encuentra también en el orden de la conducta humana, en el alma bien dispuesta y en el universo como totalidad coherente. Esta dimensión integral permite comprender la belleza como lo que agrada a la vista (*id quod visum placet*), no solo en un sentido ocular, sino como aquello que produce deleite en el conocimiento, en la contemplación y en la adhesión moral al bien. En esta línea, el acto moral bello es aquel que no presenta disonancia ni carencia; todas sus partes están ordenadas hacia el fin, y por ello puede decirse que la belleza moral refleja la armonía del alma virtuosa.

Otro aspecto relevante de la estética tomista es su carácter estructuralmente teológico. A diferencia de la estética moderna, que tiende a la autonomía del arte o al subjetivismo de lo bello, Tomás inscribe la belleza en una metafísica del ser que culmina en Dios. En Él, la belleza no es un atributo accidental, sino la esencia misma en su perfección absoluta. Dios es belleza porque es acto puro, porque su ser no está compuesto ni es susceptible de división o sombra. En Él no hay mezcla, y por eso puede decirse que su *claritas* no está superpuesta, sino que es idéntica a su esencia. En este punto, el Aquinate es claro, todo lo que posee belleza lo hace por participación; sólo Dios es bello por naturaleza.

Esta concepción onto teológica de la belleza tiene consecuencias relevantes en la comprensión del cosmos. Para Tomás, el universo entero es bello en la medida en que su multiplicidad se ordena según proporciones invisibles que reflejan la sabiduría divina. Incluso el mal, entendido, como privación tiene su lugar en esta armonía cósmica, pues permite que el bien

resplandezca con mayor claridad, como ocurre con el silencio en la música. Esta tesis, expresada con fuerza en la *Summa Contra Gentiles*, desafía toda lectura maniquea y permite vincular estética y providencia, la belleza del universo no es la suma de sus partes bellas, sino la unidad armónica que se mantiene incluso en medio de lo imperfecto.

Cabe destacar también la dimensión cristológica de la estética tomista. En su análisis de la Trinidad, Tomás afirma que el Hijo, como Verbo del Padre, es la *perfectissima pulchritudo*. En Cristo se realiza la belleza plena, no sólo porque es imagen perfecta del Padre, sino porque en Él se manifiestan todas las condiciones del *pulchrum*: proporción en su humanidad, claridad en su divinidad y perfección moral en su voluntad. Esta belleza trinitaria, que no es meramente alegórica, permite comprender por qué la contemplación de lo bello puede conducir al conocimiento del Verbo, ver lo bello, en esta perspectiva, es ya un acto intelectual que dispone al alma hacia la verdad y el bien.

Desde el punto de vista contemporáneo, resulta especialmente valioso recuperar esta concepción tomista de la belleza, en un contexto cultural donde lo bello ha sido reducido muchas veces a lo útil, lo novedoso o lo subjetivo. La propuesta tomista recuerda que lo bello posee una estructura objetiva y que su aprehensión implica tanto los sentidos como la razón. A su vez, en una época marcada por la fragmentación de lo ético, lo estético y lo religioso, el pensamiento de Tomás propone una síntesis que une el arte con la verdad y la virtud. No se trata, por supuesto, de una estética del moralismo, sino de una estética que reconoce la necesidad de la forma y la plenitud en todos los ámbitos del ser.

La estética tomista ofrece una posibilidad para reconsiderar la belleza como categoría objetiva y no reducible al gusto individual. Esto tiene implicaciones relevantes tanto para la estética teológica, donde la *claritas* puede comprenderse como transparencia hacia el misterio divino, como para la reflexión actual sobre las artes, que buscan nuevas formas de significado en un contexto cultural marcado por la fragmentación. *Integritas, proportio y claritas*

permiten repensar la creación artística como acto que manifiesta el ser y orienta al espectador hacia una comprensión más profunda de la realidad.

Finalmente, es importante subrayar que este trabajo no ha pretendido agotar la riqueza de la estética tomista, sino más bien ofrecer una estructura interpretativa que permita visibilizar sus elementos constitutivos y su relevancia actual. La belleza, en Tomás de Aquino, no es una categoría marginal, sino una clave para comprender el sentido de la creación, del conocimiento y de la vida moral. En este sentido, su pensamiento sigue ofreciendo una vía para el diálogo entre filosofía, arte, teología y cultura.

Referencias

- Aquino, S. T. (1856). *Scriptum super Sententiis* (Ed. Parmae 1856; R. Busa, transcr.; E. Alarcón, rev.).
- Aquino, S. T. (1902). *Liber contra impugnantes Dei cultum et religionem*. Sands & Co.
- Aquino, S. T. (1950). *In librum B. Dionysii De divinis nominibus expositio*. Marietti.
- Aquino, S. T. (1952). *Suma Contra Gentiles* (Vols. 1-4). Biblioteca de Autores Cristianos (BAC).
- Aquino, S. T. (1953). *Super I Epistolam B. Pauli ad Corinthios lectura*. Marietti.
- Aquino, S. T. (1965). *Expositio super Iob ad litteram*. Leonina.
- Aquino, S. T. (1989). *Suma de Teología: Primera parte de la segunda parte* (Vol. 3). Biblioteca de Autores Cristianos (BAC).
- Aquino, S. T. (1994). *Suma de Teología: Segunda parte de la segunda parte* (Vol. 4). Biblioteca de Autores Cristianos (BAC).
- Aquino, S. T. (2001). *Suma de Teología: Primera parte* (Vols. 1-2). Biblioteca de Autores Cristianos (BAC).
- Aquino, S. T. (2015). *Cuestiones disputadas sobre el mal*. Ediciones Universidad de Navarra (EUNSA).
- Aertsen, J. A. (1991). Beauty in the Middle Ages: A forgotten transcendental? *Medieval Philosophy and Theology*, 1, 68–97.
- Maritain, J. (1972). *Arte y escolástica*. Club de Lectores.
- Doctoris Angelici Divi Thomae Aquinatis*. (1895). *Opera omnia* (Vol. 5). Apud Ludovicum Vives, Bibliopolam Editorem.

La imparcialidad en la ética de Adam Smith e Immanuel Kant como fundamento de la razón pública y el liberalismo político

Impartiality in the ethics of Adam Smith and Immanuel Kant as the foundation of public reason and political liberalism

Nicolás Matías Fuentes Valdebenito

nicolas.fuentes@ucsc.cl

Universidad Católica de la Santísima Concepción

ORCID ID 0000-0002-2229-7488

Resumen: Este artículo examina comparativamente la noción de imparcialidad en Adam Smith e Immanuel Kant, con el fin de evaluar sus proyecciones normativas en la teoría política liberal contemporánea. Mientras Smith fundamenta la imparcialidad en la simpatía y la imaginación, cristalizada en la figura del espectador imparcial, Kant la concibe como un principio racional a priori, expresado en el imperativo categórico y la autonomía de la voluntad. El análisis muestra que ambos modelos, pese a sus diferencias metodológicas y epistemológicas, coinciden en exigir que la moralidad trascienda el egoísmo individual y se sitúe en una perspectiva imparcial/universal. La investigación sostiene que una síntesis entre la flexibilidad empática de Smith y la universalidad kantiana permite articular una concepción más robusta de la imparcialidad moral, capaz de responder a los desafíos de sociedades plurales. Para ello se introduce la noción de razón pública como categoría política que traduce la imparcialidad ética en legitimidad institucional: las leyes y políticas deben justificarse con argumentos accesibles a todos los ciudadanos libres e iguales. Se concluye que la integración de ambas tradiciones ofrece un fundamento normativo prometedor para el liberalismo político, en tanto conjuga sensibilidad moral y coherencia racional en la búsqueda de instituciones justas.

Palabras clave: Imparcialidad; simpatía; ética; razón pública; liberalismo político.

Abstract: This article offers a comparative examination of the notion of impartiality in Adam Smith and Immanuel Kant, with the aim of evaluating its normative projections in contemporary liberal political theory. While Smith grounds impartiality in sympathy and imagination, crystallized in the figure of the impartial spectator, Kant conceives it as an a priori rational principle, expressed in the categorical imperative and the autonomy of the will. The analysis shows that both models, despite their methodological and epistemological differences, converge in demanding that morality transcend individual egoism and adopt an impartial or universal perspective. The study argues that a synthesis between Smith's empathetic flexibility and Kantian universality enables the articulation of a more robust conception of moral impartiality, one capable of addressing the challenges of plural societies. To this end, the notion of public reason is introduced as a political category that translates ethical impartiality into institutional legitimacy: laws and policies must be justified with arguments accessible to all free and equal citizens. It is concluded that the integration of both traditions offers a promising normative foundation for liberalism, insofar as it combines moral sensitivity with rational coherence in the pursuit of just institutions.

Keywords: Impartiality; sympathy; ethics; public reason; liberalism.

Introducción

La cuestión de la imparcialidad ocupa un lugar central en la historia de la filosofía moral y política porque toca el núcleo de la justificación normativa: ¿cómo trascender los intereses particulares para alcanzar juicios y principios que puedan reclamar validez universal? Esta interrogante adquiere especial relevancia en un mundo marcado por el pluralismo, donde la legitimidad de las instituciones democráticas depende de su capacidad para sostener reglas comunes aceptables para ciudadanos diversos (Nagel, 1991). Adam Smith e

Immanuel Kant, desde tradiciones distintas, ofrecen respuestas decisivas a esta exigencia: Smith a través del espectador imparcial, sustentado en la imaginación y la simpatía; Kant mediante el imperativo categórico y la autonomía de la razón práctica. Como ha señalado White (2009): “Ambos académicos enfatizan fuertemente la imparcialidad como un elemento central de sus sistemas morales, ambos estuvieron profundamente influenciados por el pensamiento estoico y compartieron una preocupación por la dignidad humana y la libertad frente a la tiranía.” (p. 1)¹. Este trasfondo común permite comprender por qué, a pesar de sus diferentes fundamentos, ambos filósofos coincidieron en considerar la imparcialidad como condición indispensable de la moralidad.

Smith, en el marco del sentimentalismo ilustrado escocés, sostiene que la imparcialidad surge de la imaginación simpatética y del ejercicio reflexivo del espectador imparcial, que nos permite juzgar nuestras acciones como si fuéramos otro (Carrasco, 2009, p. 82). Esta figura, como ha subrayado Ben-Moshe (2023), no solo refleja las normas sociales, sino que determina “el peso relativo de al menos algunas de las razones de un agente se determina desde un punto de vista evaluativo compartido [y, en consecuencia], tienen autoridad normativa sobre las razones privadas del agente” (p. 910)². De este modo, la imparcialidad smithiana no se limita a describir el juicio moral, sino que introduce un correctivo normativo con proyección universal.

Kant, en cambio, en la tradición de la Ilustración alemana, sostiene que la imparcialidad auténtica sólo puede derivarse de un principio a priori de la razón práctica, expresado en el imperativo categórico y en la autonomía de la voluntad. La autonomía, sin embargo, ha sido objeto de lecturas equívocas que la interpretan como autodeterminación absoluta, con los riesgos sociales que ello conlleva. Como ya hemos advertido, esta interpretación resulta

¹ “both scholars strongly emphasize impartiality as a core element of their moral systems, they were both strongly influenced by Stoic thought, and they shared a concern for human dignity and freedom from tyranny” (trad. propia).

² “the relative weights of at least some of an agent’s reasons are determined from within a shared evaluative point of view [and, consequently] have normative authority” (trad. propia).

insuficiente y peligrosa, pues muchos males entienden la autonomía como la puerta al libertinaje sin responsabilidad (Bustamante-Barahona & Fuentes-Valdebenito, 2025a, p. 390). No obstante, Kant mismo evita esta reducción apelando al recurso de un espectador razonable e imparcial, que muestra que la autonomía no es licencia subjetiva sino obediencia a la ley moral que todo agente racional puede reconocer (Kant, 2017b, p. 70).

Las investigaciones recientes han mostrado que el punto de encuentro entre Smith y Kant se encuentra en el concepto de respeto. Mordacci (2023) ha sostenido que, en la teoría smithiana madura, que la aspiración más poderosa del ser humano es llegar a ser digno de respeto (p. 365), y que tanto Smith como Kant reconocen el respeto, ya sea hacia el espectador imparcial o hacia la ley moral, como un sentimiento moral fundamental que impulsa la motivación ética. Esto matiza la oposición simplista entre sentimentalismo y racionalismo: Smith no apela a meras emociones ciegas, sino a un sentimiento normativo refinado que posee una estructura racional; mientras Kant no desconoce la dimensión afectiva del respeto, aunque la entienda como efecto de la razón (Mordacci, 2023, p. 373).

En este contexto, el presente artículo no se limita a una comparación histórica entre dos tradiciones de la Ilustración, sino que busca mostrar cómo la tensión Smith–Kant permite elaborar un concepto de imparcialidad con proyección política. La hipótesis que se defiende es que una síntesis entre la imparcialidad empática de Smith y la imparcialidad racional de Kant puede articularse en el marco de la razón pública rawlsiana, entendida como el criterio mediante el cual los ciudadanos justifican las normas fundamentales en sociedades plurales. De este modo, la imparcialidad ética se convierte en imparcialidad política, sirviendo de base para la legitimidad liberal y actuando como límite negativo del Estado: las leyes y políticas sólo son legítimas si pueden justificarse con razones accesibles y recíprocamente aceptables para ciudadanos libres e iguales (Rawls, 2019).

Metodológicamente, se ha adoptado un enfoque analítico-comparativo, con tres fases de desarrollo: En primer lugar, el análisis de los fundamentos de la ética smithiana; en segundo lugar, el análisis de los

principios de la ética kantiana; y, en tercer lugar, el contraste crítico de ambas concepciones para explorar sus proyecciones en el liberalismo político. El corpus primario lo constituyen *La teoría de los sentimientos morales* y la *Fundamentación de la metafísica de las costumbres*, complementado con bibliografía especializada reciente en Smith, Kant y teoría de la razón pública. Esta metodología permite evaluar tanto la coherencia interna de cada propuesta como su capacidad para responder a desafíos normativos actuales, asegurando un análisis filosófico riguroso y con relevancia práctica.

Diferencias metodológicas: Simpatía versus Razón

La teoría ética de Adam Smith se fundamenta en una concepción empírica de la moralidad que encuentra sus raíces en la experiencia humana y en la capacidad de simpatía (Carrasco, 2014). Esta aproximación constituye lo que podríamos denominar una ética *bottom-up*, es decir, una construcción moral que emerge desde la base de las interacciones sociales, en contraposición a los sistemas éticos deductivos o apriorísticos (Fricke, 2013, p. 177). Para Smith, la moralidad no es un constructo abstracto impuesto desde arriba, sino que surge orgánicamente de las interacciones sociales y de los juicios que los individuos formulan a través de la observación e interpretación del comportamiento de otros.

Como señala Cortina (2003), este método de carácter retórico y práctico se basa en "el arte de realizar juicios probables ante situaciones individuales y concretas" (p. 27)³. En el contexto específico de la ética smithiana, este enfoque implica que los juicios morales se forman a través de la convergencia de perspectivas y experiencias compartidas, donde la simpatía actúa como mecanismo fundamental para la construcción del sentido moral. En este sentido, Fricke (2013) sostiene que: "El elemento central de su teoría

³ A diferencia del método deductivo (*top-down*), este acercamiento reconoce la importancia de las circunstancias particulares y la experiencia práctica en la formación de nuestros juicios morales, aunque, como señala Llorca (2021), es necesario trascender la simple dicotomía entre métodos deductivos e inductivos para comprender plenamente las fuentes del contenido moral.

moral es un análisis detallado del llamado «proceso simpático» que subyace a todo juicio moral” (p. 177)⁴.

En este proceso, Carrasco (2009) identifica cuatro dimensiones fundamentales de la simpatía que operan en diferentes niveles de la experiencia moral. El primer tipo es la simpatía como contagio, que representa una forma unidireccional y mecánica donde los sentimientos se transmiten involuntariamente del agente al espectador. El segundo tipo es la simpatía como identificación, también unidireccional, donde el espectador se pone imaginativamente en el lugar del agente, aunque este último no participa activamente en el proceso. El tercer tipo es la simpatía mutua, que marca un giro significativo en la teoría de Smith al introducir la bidireccionalidad: tanto el agente como el espectador participan activamente en un intercambio de perspectivas para alcanzar un punto de concordancia en los sentimientos. Finalmente, el cuarto tipo es la simpatía mutua moral, que incorpora la figura del espectador imparcial como criterio objetivo para establecer el punto de propiedad moral, superando así el potencial relativismo de la mera simpatía mutua (Carrasco, 2009, pp. 84-88).

Por su parte y en la misma línea, Alegría (2019), sostiene que el desarrollo de la capacidad simpática atraviesa tres estadios evolutivos claramente diferenciados: la mera imitación, como respuesta inicial casi instintiva; la identificación, que implica un proceso más elaborado de comprensión del otro; y finalmente, la mutua simpatía, que representa el nivel más sofisticado de interacción moral (p. 17). Esta progresión evolutiva de la capacidad simpática culmina en el desarrollo de un mecanismo moral más refinado que permite al individuo trascender la mera reciprocidad emocional para alcanzar un nivel de juicio moral más elaborado y objetivo. Este proceso de maduración moral encuentra su máxima expresión en la formación gradual del espectador imparcial.

Este concepto central en la teoría moral smithiana, el espectador imparcial, garantiza la imparcialidad en la moral. Es una construcción

⁴ “The core element of his moral theory is a close analysis of the so-called ‘sympathetic process’ which underlies every moral judgment” (Trad. propia).

imaginaria: una perspectiva objetiva e informada desde la cual evaluamos las acciones, tanto ajenas como propias, cómo lo haría un observador desinteresado y justo. En palabras de Smith:

Él es quien nos indica la corrección de la liberalidad y la deformidad de la injusticia, la propiedad de renunciar a los mayores intereses propios en aras de los intereses aún más relevantes de los demás, y la monstruosidad de perpetrar el quebrante más pequeño a otra persona con objeto de cosechar el máximo beneficio para nosotros mismos. (Smith, 2022, p. 253)

Smith señala que el ser humano no sólo busca la aprobación de los demás, sino sobre todo ser digno de aprobación, es decir, aspiramos a actuar de modo que un espectador imparcial plenamente informado apruebe nuestras motivaciones y acciones (Carrasco, 2014, p. 70).

Como señala Alegría (2019): "la estructura del juicio moral de Smith postulado en la TMS tiene dos pilares fundamentales: la noción de simpatía y la noción del espectador imparcial" (p. 17). Estos elementos no operan de manera aislada, sino que se entrelazan en un sistema dinámico que permite el desarrollo y refinamiento continuo de nuestros juicios morales. La simpatía, para Smith, constituye el puente que conecta a los individuos y facilita la formación de juicios morales: a través de la capacidad de imaginarse en la posición del otro, los seres humanos pueden desarrollar un sentido compartido de justicia y moralidad. En palabras de Weinstein (2017):

Las personas comienzan con la imaginación; esta refina la capacidad humana para razonar, lo que, combinado con los sentimientos, conduce al juicio moral. Solo después de este proceso de maduración, alguien reconoce la personalidad moral de otro. (p. 304)⁵

⁵ "People begin with the imagination, it refines the human capacity for reason which, combined with the sentiments, leads to moral judgment. Only after this maturation process does someone acknowledge another's moral personhood" (Trad. propia).

De modo similar, Alegría (2019) sostiene que: "la imaginación nos hace intercambiar posiciones con los otros y gracias a la simpatía mutua podemos llegar a realizar juicios morales y sentir placer al alcanzar tal simpatía" (pp. 17-18). Por esta razón la simpatía ocupa un rol importante en la ética smithiana donde es imposible separarla de la imaginación (Cropsey, 2020, p. 599). En consecuencia, la interacción entre simpatía e imaginación no solo fundamenta el juicio moral individual, sino que también actúa como un mecanismo esencial para la cohesión social y el fortalecimiento de los lazos comunitarios.

Como se ha visto, la concepción moral de Adam Smith, enraizada en la tradición empirista escocesa⁶, subraya que el carácter moral no es innato ni fijo, sino que se moldea a lo largo del tiempo a través de la experiencia y la interacción con los demás. Así se resalta el papel central de las emociones y la capacidad de empatizar con el otro como motores esenciales del desarrollo ético. Griswold (1999) destaca que la ética smithiana no impone reglas rígidas, sino que ofrece un modelo abierto y adaptable, capaz de responder a las particularidades y complejidades de los diferentes contextos sociales. En este sentido, la noción de simpatía y la figura del espectador imparcial funcionan como herramientas que permiten a los individuos ajustar sus juicios a las circunstancias específicas, fomentando una moralidad dinámica y en constante evolución. O en palabras de Montes (2024): "aprobamos la conducta del otro solo si existe ese equilibrio que es la simpatía mutua. Esta sintonía de sentimientos es la causa de la aprobación moral" (p. 108).

Así Carrasco (2016) afirma que cuando Griswold sostiene que la ética de Smith es fundamentalmente empírica y, por lo tanto, no puede alcanzar la universalidad, sigue operando dentro de la dicotomía moderna que separa lo empírico de lo universal. Sin embargo, Smith dismantela precisamente esa

⁶ Un aporte significativo de Carrasco (2014) en su análisis de la ética smithiana, señala que, a diferencia de Hutcheson y Hume, quienes sitúan al espectador en una posición de tercera persona, creando una estructura dicotómica con tendencia hacia el utilitarismo, Smith lo posiciona en la segunda persona, resaltando así la dimensión relacional y empática del juicio moral (p. 62).

oposición, proponiendo una visión donde la experiencia y la universalidad no se excluyen mutuamente.

La simpatía, en Smith, es un mecanismo por el cual trascendemos nuestra propia perspectiva: "no emerge tanto de la observación de la pasión como de la circunstancia que la promueve" (Smith, 2022, p. 53). Es decir, nos conmovemos no por la emoción ajena en sí, sino por entender las circunstancias que la causan. Al imaginar cómo nos sentiríamos en el lugar del otro, podemos formarnos un juicio acerca de la propiedad moral de sus acciones y emociones, saliendo así del egoísmo connatural (Benvenuto, 2019, p. 171).

Por el contrario, la ética kantiana se construye sobre una base racionalista y deontológica que redefine la imparcialidad en términos de universalidad de la ley moral y autonomía de la razón práctica. Kant afirma que "Ni en el mundo, ni, en general, tampoco fuera del mundo, es posible pensar nada que pueda considerarse como bueno sin restricción, a no ser tan sólo una buena voluntad" (Kant, 2017b, p. 69). La buena voluntad, entendida como la intención de obrar por deber moral y no por inclinaciones egoístas (imperativo hipotético), es el núcleo de la ética kantiana y el punto de partida para pensar la imparcialidad. Para Kant, una acción es moralmente valiosa no por sus consecuencias (utilitarismo) ni por la simpatía que despierte (ética smithiana), sino por la pureza de la voluntad que la ejecuta: solo cuenta la máxima por la que el agente decidió actuar, y si dicha máxima puede elevarse a ley universal (Kant, 2017b, p. 31).

Vemos aquí la imparcialidad kantiana en acto: se excluye todo privilegio personal o particular, y solo se admite como ético aquello que puede reclamarse coherentemente para cualquiera. A diferencia de Smith, Kant no apela a la imaginación emocional de ponerse en el lugar del otro, sino a la imaginación racional de un legislador universal: cada cual debe considerarse a sí mismo legislando leyes válidas para una comunidad de seres racionales en general (Kant, 2017b, p. 56). En este sentido, la perspectiva moral kantiana es la de un punto de vista universal, equivalente a la perspectiva de cualquier otro agente racional.

La imparcialidad también aparece en Kant bajo otra formulación del imperativo categórico: "Obra de tal modo que uses la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre al mismo tiempo como fin, nunca meramente como medio" (Kant, 2017b, p. 117). Esta exigencia de respeto de la dignidad de cada persona, igualmente, refleja la idea de que, desde el ángulo moral, ninguna persona puede privilegiarse o instrumentalizar a otro; todos cuentan con la misma dignidad. Kant, por tanto, internaliza la imparcialidad en la estructura misma de la ley universal y en el reconocimiento de la dignidad universal.

Un concepto central para Kant, íntimamente ligado a la imparcialidad, es la autonomía de la voluntad. Kant define la autonomía como la capacidad de la voluntad racional de darse a sí mismo la ley moral, en lugar de recibirla de fuera (heteronomía). Una voluntad autónoma es aquella que se rige únicamente por principios que ella misma, al ejercer la razón pura práctica, reconoce como universalmente válidos. En sus palabras: "Una voluntad absolutamente buena es aquella cuya máxima puede siempre tener como contenido a ella misma considerada como ley universal" (Kant, 2017b, p. 140). La autonomía kantiana garantiza la imparcialidad en la medida que el sujeto, al obedecer únicamente la ley que él mismo se da desde su razón universal, se abstrae de sus condicionamientos individuales y se coloca bajo el principio válido para cualquiera en su situación⁷.

Ahora bien, Kant mismo evita esta interpretación reduccionista de la autonomía al recurrir a una figura hipotética muy similar a la del espectador imparcial para ilustrar la prioridad de la buena voluntad. Al inicio de la Fundamentación invita a imaginar "un espectador razonable e imparcial" (Kant, 2017b, p. 70) contemplando a una persona que disfruta de una felicidad ininterrumpida, pero carece de buena voluntad; tal espectador, continúa Kant, "no podrá nunca tener satisfacción" (Kant, 2017b, p. 70) en

⁷ Frente a esto, ya hemos advertido que esta autonomía de la voluntad mal entendida como autodeterminación absoluta resulta insuficiente y peligrosa, y puede derivar en riesgos sociales de gran alcance (Bustamante-Barahona & Fuentes-Valdebenito, 2025a, p. 390). No es casual, que "muchos encuentran en la autonomía la puerta de entrada a un mundo de una libertad total, la cual en muchas ocasiones no va de la mano con la responsabilidad" (Bustamante-Barahona & Fuentes-Valdebenito, 2025b, p. 6).

ese espectáculo, y de ello concluye Kant que la buena voluntad es la condición indispensable para ser digno de la felicidad. Este ejemplo de un espectador razonable e imparcial demuestra que Kant valora la imparcialidad en el juicio moral: la aprobación moral no depende de ventajas personales ni de la parcialidad del beneficiario, sino de algo que cualquiera (desde el punto de vista razonable e imparcial) reconocería como valioso en sí: la buena voluntad. En este punto, ambos autores, Smith y Kant, convergen en imaginar un punto de vista desinteresado (sea el espectador imparcial de Smith o el espectador razonable e imparcial de Kant) como instancia validante de la moral.

No obstante, debemos reconocer que Kant desconfía de las emociones como base de la moral. Si bien reconoce que sentimos un peculiar respeto por la ley moral, un sentimiento de reverencia que la idea del deber despierta en nosotros, este sentimiento no es un punto de partida empírico sino más bien una consecuencia subjetiva del reconocimiento racional del deber.

La imparcialidad como criterio moral

El contraste metodológico entre Smith y Kant no solo refleja diferencias filosóficas fundamentales, sino que también ambas perspectivas ofrecen herramientas valiosas para abordar dilemas éticos contemporáneos. Sin embargo, sostenemos que, y como hemos ya introducido, el concepto de imparcialidad desempeña un papel central en ambas teorías, aunque su naturaleza y aplicación difieren significativamente.

Para Adam Smith, la imparcialidad se concibe como una construcción intersubjetiva, resultado de la capacidad humana de empatizar y de proyectarse en las situaciones de los demás; lo que conlleva el principio de igualdad (Carrasco, 2014, p. 80). El espectador imparcial no es una entidad externa o abstracta, sino una figura que surge a través del proceso de introspección y diálogo interno, por eso Smith lo nombra el semidios habitante de nuestro pecho (Smith, 2022, p. 246). Smith sostiene que la imparcialidad moral se construye a partir de la imaginación, permitiendo a los

individuos evaluar sus propias acciones desde la perspectiva de un tercero desinteresado, de esta forma se produce una distancia de nuestra subjetividad (Cordero, 2022, p. 164). Este proceso implica un constante ajuste emocional y una sensibilidad hacia las experiencias de otros, lo que refleja la dimensión afectiva de la ética smithiana. Por eso, ahora contrastaremos estas perspectivas destacando las principales diferencias y semejanzas, para luego examinar cómo pueden complementarse y qué aportes ofrecen conjuntamente a problemas contemporáneos en sociedades pluralistas.

En primer lugar, una gran diferencia entre ambos autores es de qué facultades humanas derivan la imparcialidad. Smith la basa en una inclinación psicológica natural: la simpatía. Según Smith, incluso el individuo más egoísta posee la capacidad de preocuparse por los demás y juzgar sus propias acciones según cómo afectarían a otros. En sus palabras:

Por más egoísta que se pueda suponer al hombre, existen evidentemente en su naturaleza algunos principios que le hacen interesarse por la suerte de otros, y hacen que la felicidad de éstos le resulten necesaria, aunque no derive de ella nada más que el placer de contemplarla. (Smith, 2022, p. 49)

La emoción y la experiencia alimentan este proceso; nuestra empatía y deseo de aprobación nos llevan espontáneamente hacia la imparcialidad encarnada en el espectador. O, en palabras de Cordero (2022), “la simpatía es su condición de posibilidad” (p. 170).

En cambio, Kant desconfía de los datos empíricos de la psicología moral. Para él, la imparcialidad auténtica debe surgir a priori de la razón. La razón práctica establece la exigencia de universalidad sin apoyarse en hechos de la naturaleza humana tampoco en todo aquello que pueda ser empírico y perteneciente a la antropología (Kant, 2017b, p. 63), sino en consideraciones lógicas (no es consistente querer una norma como universal y a la vez pretenderse excepción). Mientras Smith describe cómo juzgamos para orientarnos sobre cómo deberíamos juzgar, Kant prescribe directamente cómo deberíamos juzgar desde una perspectiva ideal de racionalidad pura. En otras palabras, la propuesta de Smith es más descriptiva y gradualmente

normativa (Campbell, 1971), anclada en la experiencia moral cotidiana, y la de Kant es plenamente normativa y trascendental, partiendo de condiciones formales de la moralidad.

En segundo lugar, tanto Smith como Kant conciben un procedimiento para elevarse por encima de la perspectiva privada, pero los mecanismos difieren. En Smith es un procedimiento imaginativo-emotivo: el individuo se proyecta en el lugar del otro y/o adopta la posición de un espectador neutral para evaluar la situación. Esto implica una especie de diálogo interno con el hombre imparcial en el pecho, donde intervienen tanto la razón común como el sentimiento decantado por la experiencia. Por ejemplo, ante un conflicto de interés personal, el agente smithiano trataría de mirarse a sí mismo como un tercero y preguntarse si un espectador imparcial aprobaría sus motivos.

En Kant, el procedimiento es estrictamente racional y universalizante: se toma la máxima de la acción y se ensaya su universalización (o se verifica si respeta la humanidad como fin), lo cual es un ejercicio mental más abstracto. Kant no nos pide imaginarnos emocionalmente en el lugar de otro individuo concreto, sino considerar si cualquier agente racional en cualquier lugar y tiempo podría seguir la misma máxima. Esta diferencia hace que la imparcialidad smithiana conserve un matiz particularista, pues el espectador imparcial valora cada caso con todos sus detalles (aunque guiado por principios generales), mientras que la imparcialidad kantiana es altamente generalista y formal: no le interesan los detalles empíricos, solo la forma lógica de la máxima.

En cuarto lugar, en la teoría de Smith, la motivación para ser imparcial proviene en buena medida del deseo de sentirse bien consigo mismo bajo la mirada del espectador imparcial. Así, Smith afirma que “Queremos pensar que somos admirables por lo que ellos también lo son. Pero para alcanzar esta satisfacción debemos transformarnos en espectadores imparciales de nuestra personalidad y conducta. Debemos procurar contemplarlos como probablemente lo harán otros” (Smith, 2022, p. 227). Smith enfatiza el anhelo humano de ser digno de aprobación: un agente virtuoso experimenta una

profunda satisfacción interna cuando sus acciones resisten el escrutinio imparcial y, a la inversa, siente remordimiento o vergüenza cuando imagina que su hombre interior desaprueba su conducta. En sus palabras:

El ser humano desea naturalmente no sólo ser amado sino ser amable, es decir, ser lo que resulta un objeto natural y apropiado para el amor. Naturalmente teme no sólo ser odiado sino ser odiable, es decir, ser lo que resulta un objeto natural y apropiado para el odio. No sólo desea la alabanza, sino el ser loable, o ser un objetivo natural y adecuado para el reproche, aunque en la práctica nadie le reproche nada. (Smith, 2022, p. 226)

Así, la ética smithiana integra motivaciones psicológicas, como el orgullo apropiado, deseo de aprobación, temor al remordimiento, para alentar la imparcialidad.

Kant, por su parte, sostiene que la única motivación verdaderamente moral es el deber mismo o, dicho de otro modo, el respeto por la ley moral. Según afirma:

Una acción hecha por deber tiene su valor moral, no en el propósito que por medio de ella se quiere alcanzar, sino en la máxima por la cual ha sido resuelta; no depende, pues, de la realidad del objeto de la acción, sino meramente del principio del querer, según el cual ha sucedido la acción, prescindiendo de todos los objetos de la facultad del desear. (Kant, 2017b, p. 78)

Ciertamente, Kant reconoce una satisfacción intelectual en obrar bien, e incluso habla del sentimiento moral de respeto; pero esas exigencias afectivas son consecuencias o acompañamientos de la acción moral, no su fundamento. El agente kantiano ideal obedece la ley universal por respeto a la ley, incluso en contra de sus inclinaciones naturales, si es preciso. La sanción interna kantiana no es tanto el remordimiento psicológico (aunque exista), sino el juicio de la propia razón: la razón dicta imperativamente lo que se debe hacer, y desobedecer implica incurrir en auto-reproche racional y degradar la propia dignidad. Por lo tanto, Smith aprovecha una motivación

psicológica existente (la búsqueda de aprobación) y la refina en conciencia imparcial, mientras Kant postula una motivación puramente racional (respeto al deber) que a su vez disciplina y limita las motivaciones sensibles.

En cuarto lugar, para Smith, la imparcialidad se aplica dentro del ámbito humano tal como lo experimentamos, con nuestras limitaciones. El espectador imparcial es comprensivo: entiende las limitaciones de la pasión humana y no exige perfección imposible. Por eso Smith distingue entre la perfección ideal que imaginaríamos en un juez totalmente informado y justo, “un semidiós dentro del pecho” (Smith, 2022, p. 246), y lo que es razonable exigir a seres humanos falibles. Su sistema permite cierto grado de variación cultural y contextual, siempre y cuando los principios básicos (justicia, benevolencia, prudencia) se conserven.

De hecho, se ha señalado un elemento de relativismo moderado en Smith: el espectador imparcial se nutre de las normas sociales vigentes, aunque también puede criticar esas normas desde ideales ilustrados de simpatía y justicia. En este sentido, el espectador imparcial no debe entenderse sólo como un reflejo de normas internalizadas, sino como una instancia que ejerce autoridad normativa sobre los juicios privados. Como explica Ben-Moshe (2023):

Sostengo que los pesos relativos de al menos algunas de las razones de un agente se determinan desde un punto de vista evaluativo compartido. Estas razones poseen autoridad normativa y constriñen las razones privadas del agente, es decir, aquellas que se determinan desde su propio punto de vista evaluativo particular. Aplico esta idea al espectador imparcial. (p. 910)⁸

⁸ “I argue that the relative weights of at least some of an agent’s reasons are determined from within a shared evaluative point of view. These reasons have normative authority over and constrain the agent’s private reasons, that is, those that are determined from within her own particular evaluative point of view. I apply this idea to the impartial spectator” (trad. propia).

Por este motivo, se ha criticado que el espectador imparcial pueda trascender dichas normas de su propia sociedad y abarcar toda la humanidad (Forman-Barzilai, 2010, pp. 86-105).

Kant, en cambio, aspira a una imparcialidad absoluta aplicable a cualquier ser racional en cualquier mundo posible. La moral kantiana no concede variación por contexto social o cultural: las mismas categorías morales, como los deberes perfectos que siempre se deben cumplir; y el mismo test de universalización valen siempre (Kant, 2017a, p. 152). Esto hace a Kant menos dispuesto a reconocer dilemas morales contextualizados o compromisos con la psicología del agente: la ley es la ley, y la razón exige su cumplimiento sin excepción⁹. De hecho, como advierte Alegría (2016), “durante la modernidad juzgar de manera imparcial equivale a juzgar como cuando se está distanciado, cuando la razón se abstrae de las particularidades y circunstancias que constituyen una situación determinada” (p. 13). No obstante, la imparcialidad kantiana no implica eliminar toda parcialidad en la acción. Según el juicio de Alegría (2021), “Para Kant el juicio moral debe ser imparcial, pero la parcialidad hacia otras personas es legítima —y en algunos casos incluso necesaria— debido a que el agente puede tener sentimientos hacia algunas personas y beneficiarlas” (Alegría, 2021, p. 42).

Y, por último, en quinto lugar, a pesar de las diferencias marcadas, Smith y Kant convergen profundamente en la noción de que la moralidad exige trascender el egoísmo individual y adoptar una perspectiva general. En este sentido, la imparcialidad smithiana anticipa la exigencia kantiana de universalidad: “Adoptar la perspectiva del espectador imparcial nos hace apreciar que nuestra propia perspectiva no es más privilegiada que —de

⁹ En esta línea, hemos mostrado que reducir la ética kantiana a la obediencia ciega resulta una interpretación profundamente errada. Como hemos advertido (Fuentes-Valdebenito & Hidalgo-Varela, 2024), Eichmann pretendió justificar sus crímenes apelando al imperativo categórico, pero una comprensión adecuada de la filosofía moral kantiana muestra que el deber no puede desligarse de la dignidad humana y de la consideración de cada persona como fin en sí mismo.

hecho, es igual a— la de otras personas” (Ben-Moshe, 2023, p. 918)¹⁰. Ambos sostienen que los sentimientos y deseos puramente privados no pueden ser la medida última de lo correcto. En Smith, la voz de la conciencia es la de un otro imparcial; en Kant, la voz de la conciencia es la de la razón universal. Incluso hay un punto específico de encuentro: el concepto de respeto. Kant eleva el respeto a la ley moral como sentimiento moral central. Smith, por su lado, aunque habla menos de respeto en términos abstractos, sí destaca el amor a lo honorable y el deseo de ser digno de respeto.

Roberto Mordacci ha mostrado que, en la teoría smithiana madura, aparece una noción de respeto hacia uno mismo y hacia los demás que enlaza con la idea de dignidad. Smith sostiene que “el deseo de convertirse en objeto adecuado de respeto es «el más fuerte de todos nuestros deseos» (TMS VI.1.3), define al espectador imparcial como un «juez respetable» (TMS III.iii.25)” (Mordacci, 2023, p. 365)¹¹. Esta formulación permite apreciar que tanto Smith como Kant convergen en considerar el respeto —ya sea hacia la ley moral en Kant o hacia los dictámenes del espectador imparcial en Smith— como un sentimiento moral fundamental que impulsa la motivación moral. En palabras del propio Mordacci, “Smith necesita que Kant encuentre una base más sólida para su distinción entre el respeto propiamente moral y la mera admiración. Kant necesita que Smith reconozca que, experiencialmente, la moralidad surge de la simpatía y el reconocimiento” (2023, p. 374)¹².

Lo anteriormente expuesto matiza la oposición simplista entre sentimentalismo y racionalismo: Smith no apela a meras emociones ciegas, sino a un sentimiento normativo refinado (respeto, aprobación imparcial) que tiene una estructura racional, mientras Kant no desconoce la dimensión

¹⁰ “Adopting the standpoint of the impartial spectator makes us appreciate that our own perspective is no more privileged than—indeed, is equal to—other people’s perspectives” (trad. propia).

¹¹ “Smith states that the desire to become the proper objects of respect is “the strongest of all our desires” (TMS VI.1.3), defines the impartial spectator as a “respectable judge” (TMS III.iii.25)” (trad. propia).

¹² “Smith needs Kant to find a firmer ground for his distinction between properly moral respect and mere admiration. Kant needs Smith to recognize that experientially, morality springs from sympathy and recognition” (trad. propia).

sentimental del respeto, aunque la entienda como resultante de la razón. En definitiva, ambos modelos buscan fundir en la moralidad la idea de universalidad con un aspecto humano: en Smith, lo humano viene dado por nuestras pasiones sociales; en Kant, por nuestra capacidad de razonar y autodeterminarnos.

Considerados conjuntamente, Smith y Kant ofrecen enfoques que pueden verse no sólo en contraste, sino también como complementarios. Smith aporta una comprensión empírica de cómo los seres humanos efectivamente desarrollan la capacidad de juicio imparcial, anclada en la psicología moral y la interacción social. Kant proporciona una formulación normativa explícita de qué exige la imparcialidad para que una norma sea justa para todos, junto con una fundamentación de la dignidad y la agencia moral. Uno podría argumentar que una ética comprehensiva de la imparcialidad podría integrar elementos de ambas: el calor humanizador de la simpatía y la firmeza estructural del principio universal. Por esta razón, Mordacci (2023) sostiene que “La búsqueda de imparcialidad de Smith es análoga a la búsqueda de universalidad de Kant, y ambos asignan a la imparcialidad y a la universalidad el papel crucial a la hora de determinar qué es apropiado o correcto” (p. 374)¹³.

De hecho, la filosofía moral y política contemporánea ha revisitado a Smith y Kant para atender a sus respectivas intuiciones: por ejemplo, la idea de un espectador imparcial resuena en teorías modernas como la del observador ideal o incluso en el método del velo de ignorancia rawlsiano; mientras que la idea kantiana de autonomía e igualdad moral es piedra angular de la concepción de derechos humanos y justicia universal.

Imparcialidad ética y razón pública en el liberalismo político contemporáneo

El concepto filosófico-político de razón pública ofrece un puente explícito entre la imparcialidad moral discutida por Smith y Kant y los problemas de

¹³ “Smith’s search for impartiality is analogous to Kant’s search for universality, and they both assign to impartiality and universality the crucial role in determining what is proper or right” (trad. propia).

legitimidad en sociedades plurales modernas. ¿En qué consiste la razón pública? En términos rawlsianos, la razón pública es el modo de razonamiento que deben emplear los ciudadanos cuando discuten cuestiones fundamentales de justicia política, de tal manera que sus argumentos puedan ser aceptables para todos los demás ciudadanos libres e iguales, sin apelar a doctrinas sectarias o visiones particulares del bien (Chandler, 2025, p. 84). En palabras de Rawls (2019): “la razón pública es característica de un pueblo democrático: es la razón de sus ciudadanos, de quienes comparten una posición de igual ciudadanía” (p. 247). Tal como ha sido interpretado, “Para Rawls, la razón se vuelve pública cuando se generaliza, se anonimiza, se descontextualiza y se universaliza” (Boutieri et al., 2025, p. 4)¹⁴. Esta formulación sintetiza la exigencia de que las razones políticas se expresen en un lenguaje accesible a todos los ciudadanos, sin depender de doctrinas particulares o posiciones identitarias. Dicho de forma sencilla, la razón pública exige imparcialidad en la justificación de las leyes y políticas básicas: uno debe dar razones que cualquiera (cualquier ciudadano razonable, independientemente de su religión, ideología comprensiva o intereses privados) podría aceptar en principio.

El requisito de que el poder sea legítimo solo cuando es justificable públicamente se articula en el principio de reciprocidad: cada ciudadano, al participar en la deliberación pública, avance razones que, a su entender sincero, las demás personas libres e iguales podrían aceptar sin menoscabo de su dignidad. La reciprocidad, así entendida, convierte la interacción política en un ejercicio de mutuo reconocimiento: no se trata de persuadir mediante la fuerza retórica ni de imponer convicciones privadas, sino de proponer bases normativas susceptibles de ser compartidas por interlocutores que mantienen doctrinas comprensivas diversas (Macedo, 1990, p. 78). Por el contrario, para Rawls una concepción política es irrazonable si no satisface el principio de reciprocidad (Freeman, 2016, p. 361).

¹⁴ “For Rawls, reason becomes public when it is generalized, anonymized, decontextualized, and universalized” (trad. propia).

Según nuestra visión, esto impone un límite negativo al Estado, en el sentido de que las autoridades públicas no deben fundamentar las leyes en motivos que sean accesibles solo para un grupo parcial de la sociedad, ya que esto crearía ciudadanos de primera y segunda categoría (Cortina, 2014). La coerción estatal sólo sería legítima cuando se apoya en razones públicas, es decir, razones imparciales, compartibles en el espacio común.

Se puede apreciar cómo esta idea es afín a los principios de Smith y Kant. Por un lado, la razón pública se parece a un espectador imparcial a nivel social: en una democracia plural, cada ciudadano al deliberar debería tratar de adoptar el punto de vista del ciudadano imparcial, abstraído de sus propias creencias particulares, para formular argumentos basados en valores políticos comunes (por ejemplo, la libertad y la igualdad) que los demás puedan evaluar y compartir (Villavicencio-Miranda, 2009, p. 535). Esta exigencia recuerda la invitación smithiana a ponerse en el lugar de otros antes de juzgar o decidir, ampliada aquí al ámbito público.

Por otro lado, la razón pública refleja el imperativo kantiano de universalización aplicado a la legislación: una norma política es legítima sólo si puede reclamarse con carácter universal dentro de la ciudadanía, esto es, sin tratar a nadie como mera herramienta de los fines de otros. En palabras de Villavicencio-Miranda:

Lanzados en un mundo en el que nos encontramos divididos por profundas diferencias respecto de nuestras doctrinas comprensivas, muchas de ellas irreconciliables, la sensata necesidad de alcanzar los beneficios de la cooperación social conduce a escindir las razones que valen para mí de las que pueden valer para todos, incluso para aquellos que piensan muy distinto. (Villavicencio-Miranda, 2009, p. 534)

Rawls mismo reconoció la influencia kantiana en su pensamiento: el carácter público de la razón se vincula a la idea de que los ciudadanos son co-legisladores autónomos que deben darse leyes mutuamente aceptables, análogamente a como la voluntad moral kantiana se da leyes universales.

Cabe señalar que la razón pública actúa como límite negativo: marca qué no puede hacer legítimamente el Estado liberal (por ejemplo, no puede imponer dogmas religiosos o ideológicos particulares como base de las leyes, ya que estos no cuentan como razones públicas compartidas). Este límite protege la imparcialidad del poder político en un contexto de diversidad: ningún grupo mayoritario puede justificar coerción estatal únicamente con sus creencias privadas. De este modo, la imparcialidad normativa que Smith y Kant exigían en moral personal se traslada a las instituciones: las políticas deben ser evaluadas desde una perspectiva imparcial de la ciudadanía. Cuando los individuos formulan políticas públicas poniendo entre paréntesis sus lealtades parciales y buscando fundamentos comunes, están operando bajo el ideal de la razón pública.

La síntesis ética que hemos delineado entre Smith y Kant — combinando la sensibilidad hacia la condición humana que aporta la simpatía con la firme exigencia de universalidad que aporta el imperativo categórico— puede fortalecer la noción de razón pública. Un sistema político justo requeriría instituciones que encarnen la imparcialidad tanto en contenido como en procedimiento: por contenido, las leyes deben respetar por igual la dignidad e intereses básicos de todos (visión kantiana), y por procedimiento, deben surgir de una deliberación pública empática donde se consideren las distintas perspectivas sociales (visión smithiana). En contextos plurales, la legitimidad institucional aumenta si los ciudadanos perciben que las decisiones se toman con un espíritu de espectador imparcial social, y no obedeciendo a facciones o dogmas particulares. Como sugiere Villavicencio-Miranda, siguiendo a Rawls:

Cada uno de nosotros ingresa [en la sociedad] sin poder alejarse de ella, pero que nos demanda igualmente un mecanismo a través del cual podamos explicarnos mutuamente el fundamento de nuestras acciones públicas, de modo que todos podamos unos a otros el fundamento de nuestras acciones públicas, de modo tal que cada cual pueda esperar razonablemente que los demás aceptarán ese fundamentos. Esa es la cuestión a la cual debe enfrentarse la razón pública. (Villavicencio-Miranda, 2009, pp. 534-535)

Ese mecanismo es la razón pública, que no es sino la imparcialidad moral traducida al lenguaje de la deliberación política. Si los individuos han cultivado en sí mismos la actitud moral imparcial, ya sea bajo la forma de la consciencia simpatética smithiana o del respeto autónomo kantiano, estarán mejor preparados para cumplir con el deber cívico de justificar y demandar leyes imparciales.

Así, la reflexión ética sobre Smith y Kant no es un ejercicio meramente histórico o teórico, sino que tiene eco en problemas contemporáneos de justicia: la tensión entre pluralismo y cohesión social, el tipo de neutralidad que debe (o no) tener el Estado, y la forma en que la ciudadanía puede sobreponerse a divisiones religiosas o ideológicas para coexistir bajo principios comunes imparciales. Por esta razón:

En el liberalismo político, para participar en la razón pública, es necesario acceder a ella mediante la práctica de la autoabstracción. Esta restricción exige que las personas formulen afirmaciones y argumentos de forma que otros puedan comprenderlos y potencialmente aceptarlos, despojándose de cualquier característica parroquial de su cosmovisión. (Boutieri et al., 2025, p. 6)¹⁵

La imparcialidad en el ámbito moral individual se proyecta en un ideal de imparcialidad política. En última instancia, podríamos decir que la ética proporciona las condiciones de legitimidad moral que luego la teoría política eleva a condiciones de legitimidad institucional.

Por último, Smith nos recuerda que detrás de las leyes hay seres humanos con sentimientos, cuya aprobación es necesaria; Kant nos recuerda que esas leyes deben responder a principios racionales que cualquiera pueda reconocer. En este sentido, sostiene Alegría (2011), “para juzgar imparcialmente se requiere respetar al otro por lo que el otro es y lo que ha

¹⁵ “In political liberalism, in order to partake in public reason, one must enter the public sphere through the practice of self-abstraction. This constraint requires people to make claims and arguments in such a manner that others can understand and potentially accept, and that are stripped of any parochial characteristics of worldview” (trad. propia).

hecho, lo que a su vez requiere que se lo pueda mirar simultáneamente desde distintos puntos de vista para poder «ver al otro» tal como verdaderamente es, y reconocer sus demandas” (pp. 16-17). La razón pública armoniza estas exigencias: es un estándar que apela tanto a la persuasión racional universal como a la comprensión mutua entre distintos, invitando a los ciudadanos a razonar como espectadores imparciales de las políticas que adoptan, guiados por consideraciones de justicia que trascienden su situación personal.

Conclusiones

En conclusión, la comparación de las éticas de Adam Smith e Immanuel Kant en torno a la imparcialidad revela una rica complementariedad tras su contraste inicial, aportando lecciones valiosas para la filosofía moral y política. Recapitulando, Smith concibe la imparcialidad moral arraigada en la naturaleza humana social: mediante la simpatía y la figura del espectador imparcial, la moralidad se nutre de la capacidad de los individuos para salir de sí mismos emocional e imaginativamente, buscando la aprobación de un juez interno honesto e igual para todos. Kant, por su lado, erige la imparcialidad sobre la autonomía racional: la moralidad exige salir de nuestros condicionamientos subjetivos y someterse solo a aquellas leyes que cualquiera podría darse, tratando a cada persona con igual dignidad. Las diferencias son manifiestas, sentimiento vs. razón, inducción empírica vs. deducción apriorística, flexibilidad contextual vs. rigidez universal, etc., pero ambos sistemas convergen en el ideal normativo de una perspectiva moral universal y desinteresada. Cada uno, a su manera, busca solucionar el problema fundamental de la ética: cómo superar la parcialidad del punto de vista individual para alcanzar juicios y decisiones que merezcan aceptación general.

Al llevar estas ideas al terreno político, se ha sostenido que la síntesis ética aquí propuesta (integrando elementos de Smith y Kant) tiene importantes consecuencias para la teoría de la justicia y la legitimidad institucional en democracias liberales. Una ética imparcial que combine la empatía informada con la universalidad del principio puede generar ciudadanos más conscientes de las razones de los otros y más comprometidos con principios de justicia. En un contexto de diversidad cultural y moral, la

imparcialidad no es un lujo, sino una condición de posibilidad para la coexistencia pacífica: solo si somos capaces de justificar nuestras demandas en términos imparciales (razón pública) podremos sostener instituciones legítimas que todos reconozcan como justas (Rawls, 2019). Tanto Smith como Kant ofrecerían su aprobación a esta idea, aunque desde enfoques distintos: Smith enfatizaría la importancia de la mirada del otro, de considerar los intereses y sentimientos de todos los afectados (lo cual se refleja en la necesidad de una deliberación pública inclusiva), mientras que Kant subrayaría la necesidad de principios universales inviolables (reflejados en los derechos fundamentales y la igual consideración de cada persona ante la ley).

La articulación de imparcialidad moral e imparcialidad política sugiere que cultivar individuos virtuosos —con conciencia imparcial— favorece la construcción de estructuras sociales justas, y recíprocamente, instituciones imparciales educan a los individuos en la virtud de la justicia, lo que en términos de Boutieri et al. (2025), sería el imperativo de la autoabstracción. La ética comparada de Smith y Kant nos enseña que la imparcialidad tiene múltiples facetas complementarias: una sentimental-social y una racional-formal. Lejos de ser excluyentes, estas dimensiones se potencian mutuamente en la búsqueda de una sociedad bien ordenada. Por ejemplo, principios de justicia formulados imparcialmente (a la manera kantiana) requieren de la empatía y la comprensión del contexto humano (a la manera smithiana) para aplicarse correctamente y para ganar adhesión entre las personas reales; a su vez, la empatía sin un marco de principios universales puede degenerar en favoritismo o sentimentalismo ineficaz. Así, la combinación de ambos enfoques proporciona un modelo más robusto de moralidad pública, donde las leyes justas son aquellas que un espectador imparcial aprobaría y que satisfacen un principio de universalización racional.

Finalmente, este análisis renovado abre proyecciones teórico-políticas: sugiere un fundamento ético para los derechos humanos y los consensos básicos de la democracia liberal. En este sentido, “se puede afirmar que el consenso no es el fin de la política, sino su condición de posibilidad: una forma de cooperación política que respeta la pluralidad, siempre que se manifieste de forma razonable” (Carbajal et al., 2025, p. 31). Los ecos de

Smith y Kant resuenan, por ejemplo, en la idea de que la justicia debe equilibrar consideraciones de equidad concreta (atender a las necesidades y voces de todos, especialmente de los menos favorecidos, lo que requiere simpatía e imparcialidad en la mirada) con reglas imparciales generales (igualdad ante la ley, normas universales de derechos y deberes). La legitimidad de las instituciones se refuerza cuando se perciben como imparciales tanto en procedimiento como en contenido, encarnando ese juez ideal que ni se inclina por poderosos ni excluye a minorías, sino que gobierna desde el pecho de cada ciudadano razonable, tal como lo imaginaron Smith y Kant cada uno a su modo. En conclusión, lograr una síntesis entre la ética de la simpatía imparcial de Smith y la ética del deber autónomo de Kant proporciona una plataforma normativa sólida para abordar los desafíos de la justicia y la legitimidad en un mundo plural, recordándonos que la verdadera imparcialidad es a la vez humana y racional, emotiva y universal, y que en ella descansa tanto la virtud personal como la paz social duradera.

Referencias

- Alegría, D. (2019). Juicios intersubjetivos. Algunas similitudes en Kant, Smith y Arendt. *Revista de Filosofía UCSC*, 17(2), 11–27. <https://doi.org/10.21703/2735-6353.2018.17.02.01>
- Alegría, D. (2021). ¿Puede un kantiano ser parcial con sus seres queridos? Una discusión frente a objeciones recurrentes. *Signos Filosóficos*, 23(46), 38–59. <https://signosfilosoficos.izt.uam.mx/index.php/SF/article/view/738>
- Alegría, D. (2023). Críticas a la exigencia de imparcialidad en la moral y una propuesta de reinterpretación. *Revista de Filosofía UCSC*, 15(2). <https://revistas.ucsc.cl/index.php/revistafilosofia/article/view/2330>
- Ben-Moshe, N. (2023). An Adam Smithian account of humanity. *Ergo*, 10(32). <https://doi.org/10.3998/ergo.4662>
- Benvenuto, R. M. (2019). Los límites éticos del mercado. Hegel- Smith y la crítica de la economía política. En H. Borisonik, F. Ludueña Romandini & J. Acerbi (Eds.), *Detrás del espectador imparcial. Ensayos en torno de Adam Smith* (pp. 167–192). Instituto de Investigaciones Gino Germani.
- Boutieri, C., Everett, S. S., & Weiss, E. (2025). Introduction: Beyond public reason. *Journal of the Royal Anthropological Institute*. <https://doi.org/10.1111/1467-9655.14318>

- Bustamante-Barahona, P., & Fuentes-Valdebenito, N. (2025a). La autonomía en la fertilidad: análisis personalista de los derechos sexuales y reproductivos desde una revisión histórico-bioética. *Ius Humani. Revista De Derecho*, 14(2), 375–404. <https://doi.org/10.31207/ih.v14i2.439>
- Bustamante-Barahona, P., & Fuentes-Valdebenito, N. (2025b). La propuesta cristiana de los cuidados paliativos como respuesta a la eutanasia y al dilema de la autonomía absoluta. *Cuadernos De Teología*, 17, e6670. <https://doi.org/10.22199/issn.0719-8175-6670>
- Campbell, T. (1971). *Adam Smith's science of morals*. Routledge.
- Carrasco, A. (2009). De Hutcheson a Smith: un sentimentalismo 'sofisticado'. *Revista de Filosofía*, 65, 81–96. <https://doi.org/10.4067/S0718-43602009000100005>
- Carrasco, A. (2014). Reinterpretación del espectador imparcial: impersonalidad utilitarista o respeto a la dignidad. *Crítica: Revista Hispanoamericana de Filosofía*, 46(137), 61–84. <http://www.jstor.org/stable/43854521>
- Carrasco, A. (2016). La ética de Adam Smith: Conciliando paradigmas, una propuesta olvidada. *Trans/Form/Ação*, 39(3), 23–28. <https://doi.org/10.1590/S0101-31732016000300003>
- Chandler, D. (2025). *Libres e iguales. Un manifiesto por una sociedad justa*. Paidós.
- Cordero, T. (2022). La simpatía como mecanismo de juicio moral en la ética de Adam Smith. *Colloquia. Revista académica de Pensamiento y Cultura*, 9, 159–173. <https://doi.org/10.31207/colloquia.v9i0.131>
- Cortina, A. (2003). El quehacer público de la ética aplicada: ética cívica transnacional. En A. Cortina y D. García-Mazá (Eds.), *Razón pública y éticas aplicadas. Los caminos de la razón práctica en una sociedad pluralista* (pp. 13–44). Tecnos.
- Cortina, A. (2014). *Alianza y Contrato. Política, ética y religión*. Trotta.
- Cropsey, J. (2020). Adam Smith. En L. Strauss & J. Cropsey, *Historia de la filosofía política* (pp. 597–618). Fondo de Cultura Económica.
- Forman-Barzilai, F. (2010). *Adam Smith and the circles of sympathy*. Cambridge University Press.
- Freeman, S. (2016). *Rawls*. Fondo de Cultura Económica.
- Fricke, C. (2013). The sympathetic process and the origin and function of conscience. En C. Berry, M. P. Paganelli, & C. Smith (Eds.), *The Oxford handbook of Adam Smith* (pp. 177–200). Oxford University Press.
- Fuentes-Valdebenito, N. M., & Hidalgo-Varela, L. S. (2024). La deontología en la filosofía moral kantiana: El caso de Adolf Eichmann en la Solución Final. *Mutatis Mutandis: Revista Internacional De Filosofía*, 11, 1–12. <https://doi.org/10.69967/07194773.v11i.499>
- Griswold, C. (1999). *Adam Smith and the virtues of enlightenment*. Cambridge University Press.
- Kant, I. (2017a). *Crítica de la razón práctica*. Tecnos.
- Kant, I. (2017b). *Fundamentación de la metafísica de las costumbres*. Tecnos.

- Llorca, J. (2021). *Análisis y reconstrucción de la idea filosófica de ética aplicada desde la hermenéutica crítica de Adela Cortina* [Tesis de máster, Universidad Complutense de Madrid]. <https://hdl.handle.net/20.500.14352/5153>
- Macedo, S. (1990). *Liberal virtues: Citizenship, virtue and community in liberal constitutionalism*. Clarendon Press.
- Montes, L. (2024). *Adam Smith. El filósofo economista*. Fondo de Cultura Económica / Centro de Estudios Públicos.
- Mordacci, R. (2023). From sympathy to respect: Smith and Kant on moral motivation. *History of Philosophy Quarterly*, 40(4), 359–378. <https://doi.org/10.5406/21521026.40.4.05>
- Nagel, T. (1991). *Equality and partiality*. Oxford University Press.
- Rawls, J. (2019). *El liberalismo político*. Crítica.
- Smith, A. (2022). *La teoría de los sentimientos morales*. Alianza.
- Villavicencio-Miranda, L. (2009). Algunas críticas a la idea de razón pública rawlsiana. *Revista de derecho (Valparaíso)*, (32), 533–557. <https://doi.org/10.4067/S0718-68512009000100015>
- Weinstein, J. R. (2017). Kant and Smith on imagination, reason, and personhood. En E. Robinson & C. W. Surprenant (Eds.), *Kant and the Scottish Enlightenment* (pp. 304–325). Routledge.
- White, M. D. (2009). Adam Smith and Immanuel Kant: On markets, duties, and moral sentiments. *Forum for Social Economics*. <https://ssrn.com/abstract=1318605>

La Inteligencia Artificial en la formulación y personalización de políticas públicas: Entre la eficiencia, la equidad y el riesgo democrático

Artificial Intelligence in the formulation and customization of public policies: Between efficiency, equity, and democratic risk

Dayanara Noelly Solano Carpio

dnsolano@utpl.edu.ec

Universidad Técnica Particular de Loja

ORCID ID 0009-0003-4871-1530

Boris Raúl Ochoa Ordóñez

brochoa@utpl.edu.ec

Universidad Técnica Particular de Loja

ORCID ID 0000-0003-4928-0915

Resumen: La incorporación de sistemas de inteligencia artificial (IA) en el ciclo de las políticas públicas está transformando de manera sustantiva la manera en que los Estados diagnostican problemas, diseñan intervenciones, asignan recursos y evalúan resultados. Este artículo examina críticamente cómo la IA reconfigura las relaciones entre eficiencia administrativa, equidad distributiva, transparencia institucional y legitimidad democrática, y propone un marco operativo para su adopción responsable en contextos públicos. Este análisis se sustenta en la pregunta ¿bajo qué condiciones los beneficios productos del uso de la IA justifican los riesgos que representan para la gobernanza democrática contemporánea? Para responder esta pregunta, esta investigación elabora un análisis conceptual y metodológico con base en cuatro pilares: eficiencia, equidad, transparencia y legitimidad. En cada uno de estos se empieza con una definición operativa, pasando por señales verificables que

permiten evaluar su nivel de cumplimiento y, por último, definir los límites normativos que deberían trazar el camino de su aplicación. De igual manera, la investigación marca una diferencia entre personalización política y personalización administrativa, evidenciando que la personalización administrativa mejora la concentración y provisión de servicios. No obstante, también puede incrementar las desigualdades si es que existen sesgos en los datos o una baja supervisión humana. El artículo plantea una secuencia de gobernanza antes, durante y después de la implementación que enlaza evaluación de impacto algorítmico, dirección humana en puntos clave, auditoría externa periódica, trazabilidad, publicidad de fundamentos y mecanismos prácticos de apelación. Partiendo de casos internacionales de salud, protección social, educación, se identifican requisitos que autorizan y supuestos de veto para asegurar que la eficiencia no se priorice sobre la justicia social, la no discriminación y la correcta rendición de cuentas. Esta investigación culmina señalando que la IA puede ayudar a fortalecer la capacidad del estado, pero solo si su implementación se realiza en una arquitectura institucional que resguarde la participación ciudadana, el control democrático y la transparencia. La IA no sustituye el juicio humano ni el proceso deliberativo; lo complementa cuando existen salvaguardas claras, estándares verificables y una supervisión continua orientada al bienestar público.

Palabras clave: Inteligencia artificial; política pública; toma de decisiones algorítmica; gobernanza digital.

Abstract: The integration of artificial intelligence (AI) systems into the public policy cycle is substantially transforming how states diagnose problems, design interventions, allocate resources, and evaluate outcomes. This article critically examines how AI reconfigures the relationships between administrative efficiency, distributive equity, institutional transparency, and democratic legitimacy, and proposes an operational framework for its responsible adoption in public contexts. This analysis is based on the question: under what conditions do the benefits of AI justify the risks it poses to contemporary democratic governance? To answer this question, this research develops a conceptual and methodological analysis based on four pillars: efficiency, equity, transparency, and legitimacy. Each pillar begins with an operational definition, moves on to verifiable indicators that allow for the assessment of its level of compliance, and finally defines the normative limits that should guide its application. Similarly, the research distinguishes between political personalization and administrative personalization, demonstrating that administrative personalization improves the concentration and provision of services.

However, it can also exacerbate inequalities if there are biases in the data or weak human oversight. The article proposes a governance sequence before, during, and after implementation that links algorithmic impact assessment, human guidance at key points, periodic external audits, traceability, disclosure of rationale, and practical appeals mechanisms. Drawing on international cases in health, social protection, and education, it identifies enabling requirements and veto assumptions to ensure that efficiency is not prioritized over social justice, non-discrimination, and proper accountability. This research concludes by noting that AI can help strengthen state capacity, but only if its implementation takes place within an institutional framework that safeguards citizen participation, democratic control, and transparency. AI does not replace human judgment or the deliberative process; it complements them when there are clear safeguards, verifiable standards, and continuous oversight focused on the public good.

Keywords: Artificial intelligence; public policy; algorithmic decision-making; digital governance.

Introducción

La implementación de sistemas de inteligencia artificial (IA) en la gestión públicas se ha convertido actualmente en uno de los focos de atención en el debate académico y político sobre el futuro del Estado, la democracia y la justicia social. Su uso en las diferentes etapas del ciclo de creación, implementación y evaluación de políticas públicas augura incrementar la eficiencia administrativa, optimizar la focalización de recursos y aumentar la capacidad del Estado para prever riesgos y organizar intervenciones. No obstante los beneficios se opacan por temas como desigualdad estructural, sesgos logarítmicos, toma de decisiones automáticas y sin transparencia, dependencia tecnológica y débil control ciudadano.

Estudios recientes están de acuerdo en que estos problemas son parte fundamental de la manera en la que los sistemas algorítmicos reorganizan la

relación entre el Estado, los ciudadanos y los procesos de elección democrática (Eubanks, 2018; Yeung, 2018).

Aún con el creciente interés en la implementación de la IA en los gobiernos nacionales y locales, todavía perduran brechas analíticas y normativas que obstaculizan valorar de forma integral su huella en la capacidad estatal y en la legitimidad institucional. Parte importante de los estudios se enfocan en aplicaciones sectoriales como la salud, educación, protección social o en diagnósticos generales sobre riesgos éticos, sin elaborar marcos operativos que permitan a las administraciones públicas disponer cuándo, cómo y en qué contextos la IA puede utilizarse de forma responsable (Kroll, 2020; Shneiderman, 2022). Esta ausencia de criterios aplicables genera dos riesgos simultáneos: por una parte, procesos de automatización precipitados que profundizan desigualdades; por otra, restricciones excesivas que limitan la innovación pública aun cuando podrían implementarse salvaguardas adecuadas.

Este artículo parte de una pregunta central: ¿en qué condiciones los beneficios asociados al uso de IA justifican los riesgos que introduce para la gobernanza democrática contemporánea? Esta interrogante no pretende confrontar eficiencia y democracia como extremos incompatibles, sino más bien identificar las condiciones en las que la IA puede fortalecer la capacidad estatal sin poner en tela de juicio los valores fundamentales de equidad, transparencia y legitimidad. En consecuencia, se aplica un enfoque conceptual-analítico estructurado en cuatro pilares: eficiencia, equidad, transparencia y legitimidad.

Cada pilar se despliega desde una definición operativa, los mecanismos que explican como la IA afecta dicho ámbito, las señales verificables que permiten evaluar su nivel de cumplimiento, y los límites normativos que deberían ser los ejes sobre su utilización.

La principal contribución de este artículo es convertir un análisis general sobre ventajas y riesgos en un marco evaluativo y operativo aplicable por instituciones públicas. En contraste con las aproximaciones basadas en

listados de riesgos o en manuales de buenas prácticas, este estudio propone un enfoque secuencial que enlaza lo que debe suceder antes, durante y después del despliegue de sistemas algorítmicos. Esto es: (a) la realización de evaluaciones de impacto algorítmico previas a la implementación, (b) establecimiento de umbrales mínimos para la automatización de procesos, (c) la obligatoriedad de supervisión humana en puntos clave, (d) mecanismos fuertes de trazabilidad, auditoría externa y apelación ciudadana una vez que la IA se encuentra funcionando (Kaminski & Malgieri, 2021; Raji et al., 2020).

La introducción conceptualiza la diferencia entre personalización política que tramita procesos de comunicación, liderazgo y representación, y la personalización administrativa que define las formas en las que el Estado modifica los recursos y servicios en función de las características individuales o grupales de la población.

Esta diferenciación no es irrelevante ya que confundirlas deriva en diagnóstico equivocados y sesgos en la interpretación. Este artículo se enfoca exclusivamente en la personalización administrativa y analiza sus efectos sobre la equidad, eficiencia y legitimidad.

Por último, en lo que al alcance del texto tiene relación, este no es un estudio empírico ni busca evaluar un caso nacional o local específico, sino construir un marco conceptual anclado en bibliografía interdisciplinaria reciente (2020 – 2025) y una evidencia comparada sobre la implementación de IA en servicios públicos. Se busca establecer criterios operativos aplicables a instituciones de diversos niveles de gobierno interesadas en incorporar la IA de forma responsable, transparente y acorde a los principios democráticos.

En síntesis la introducción establece la relevancia del debate define el enfoque analítico, presenta la pregunta de investigación y destaca de forma breve la contribución del artículo en el ámbito de la gobernanza algorítmica y las políticas públicas.

Metodología

Esta investigación adopta un enfoque metodológico mixto que combina un análisis conceptual-analítico con una revisión narrativa estructurada de literatura interdisciplinaria reciente. Este diseño metodológico permite avanzar desde una comprensión teórica de los mecanismos mediante los cuales la inteligencia artificial (IA) transforma la acción pública hacia la identificación de patrones, tensiones y criterios aplicables en contextos gubernamentales concretos. La mezcla de ambos enfoques cumple con la necesidad de articular dentro de un mismo marco, los fundamentos normativos de la gobernanza algorítmica y la evidencia empírica reciente sobre el uso de la IA en sectores como la salud, educación, administración pública digital y protección social.

La parte conceptual-analítica empieza con la revisión de teoría relacionada con la regulación algorítmica, equidad, transparencia y legitimidad democrática. Busca definir de manera operativa los conceptos de eficiencia, equidad, transparencia y legitimidad.

Estos conceptos están formados por criterios de validez, indicadores para medir su efectividad y características causales que permiten analizar la forma en que la IA interviene en la dinámica estatal.

Esta investigación tomó como referente literatura especializada en la ética en el uso de la IA, políticas públicas y gobernanza digital, permitiendo la convergencia entre la teoría, los riesgos documentados y evaluaciones institucionales.

El segundo componente de la metodología hace referencia a la revisión narrativa estructurada de literatura publicada entre 2020 y 2025. La delimitación de estos años hace referencia al rápido avance que han tenido las tecnologías de aprendizaje automático y al surgimiento de marcos regulatorios recientes que rediseñan los estándares de gobernanza algorítmica.

Se incluyeron en el análisis, informes técnicos de organismos multilaterales, artículos científicos, documentos de políticas públicas y

estudios comparados y enfocados en el uso de la IA en gobiernos nacionales y locales. Para la revisión bibliográfica se consultó bases de datos académicas como Google Scholar, Scopus, entre otras y se complementó con información encontrada en repositorios especializados en gobernanza digital y políticas públicas.

Se eligió las fuentes de información priorizando los estudios referentes a la aplicación de la IA en funciones del Estado, investigaciones que evalúan el impacto de la IA en los cuatro pilares antes mencionados, documentos que se enfocan en mecanismos de auditoría y control y bibliografía que define la ética en el uso de IA.

No se usaron documentos que eran puramente técnicos ni que no guardaban relación con la política pública. También se excluyeron estudios en los que se analiza el uso de la IA con fines de lucro y, por tema de actualidad, tampoco artículos publicados antes del 2019.

En principio se recolectó información enfocada en los cuatro pilares (eficiencia, equidad, transparencia y legitimidad democrática), y luego se buscó información relacionada con formas de dar a entender los procesos tecnológicos detrás de la IA, maneras de procurar la protección institucional, auditorías y control ciudadano.

Este enfoque mixto permitió diseñar una secuencia operativa de gobernanza algorítmica construida mediante la comparación entre nuevos marcos normativos y evidencia empírica sobre implementaciones exitosas y fallidas. Esta secuencia busca transformar principios abstractos en pasos verificables antes, durante y después del despliegue de sistemas algorítmicos en el sector público.

Finalmente, se reconocen las limitaciones metodológicas del estudio ya que no corresponde a una revisión sistemática exhaustiva ni a estudios de campo, sino más bien a una aproximación híbrida dirigida a la construcción de un marco conceptual robusto y operativo. Aun así, este enfoque permite detectar tendencias consistentes en la literatura reciente, la reconstrucción de

mecanismos de impacto y la elaboración de criterios prácticos para el fortalecimiento de la gobernanza democrática en contextos de rápida digitalización.

Marco teórico y ejes analíticos

El análisis de la inteligencia artificial (IA) en el sector público requiere un marco teórico capaz de articular dimensiones administrativas, tecnológicas y normativas. Este artículo acoge una visión que entiende a la IA como un conjunto de sistemas algorítmicos que transfiguran la forma en que los Estados analizan la información, coordinan intervenciones y reparten los bienes públicos. Empero, estas transfiguraciones no se limitan al ámbito técnico, también reajustan relaciones de poder, patrones de equidad y la misma estructura de la legitimidad democrática. Desde esta perspectiva, los cuatro pilares de análisis: eficiencia, equidad, transparencia y legitimidad, constituyen un marco para entender como la forma en que la IA incide en las principales funciones del Estado.

Así entonces, la eficiencia es uno de los principales argumentos para la adopción de la IA en los gobiernos gracias a su capacidad de transformar recursos en resultados dado que la automatización de tareas, reducción de errores y optimización de flujos de decisión prometen incrementar la productividad estatal. Sin embargo, la eficiencia carece de neutralidad puesto que depende de cómo se definen los objetivos y de que tan equitativa sea la distribución de beneficios.

Algunos de los principales indicadores de eficiencia son: tiempo de procesamiento, reducción de costos administrativos, exactitud en la gestión de datos y capacidad de predicción. No obstante, estos beneficios deben delimitarse ya que la eficiencia no debe justificar la automatización en la toma de decisiones que afecten derechos sin mecanismos claros de supervisión humana y espacios para la apelación.

La regla operativa derivada establece que solo puede adoptarse IA en procesos críticos cuando existan métricas verificables de desempeño y salvaguardas activas que garanticen la corrección de errores.

La equidad examina cómo la IA distribuye beneficios y riesgos entre toda la población. Los modelos algorítmicos funcionan a partir de datos que, muchas veces, reflejan desigualdades estructurales.

La literatura ha demostrado que los sistemas de puntuación, clasificación o predicción suelen generar patrones de exclusión cuando los datos no son representativos, sobre todo cuando a las poblaciones vulnerables hacen referencia (Barocas & Hardt, 2020; Mehrabi et al., 2022). Los componentes que pueden explicar estos efectos pueden ser sesgos de muestreo, decisiones de diseño que priorizan la exactitud sobre la justicia distributiva y la falta de supervisión humana.

Los principales indicadores de equidad son las métricas distributivas y la falta de disparidades no justificadas. Desde una perspectiva normativa se dispone que ningún sistema automatizado debe implementarse sin una evaluación previa de impacto algorítmico ni sin mecanismos de corrección permanente.

Desde el ámbito administrativo, el uso de la IA es aceptable cuando ayuda a reducir desigualdades, o como mínimo no las incrementa y cuando se tiene salvaguardas para identificar y corregir efectos adversos.

La transparencia se refiere a la capacidad institucional de comprender cómo y por qué un sistema algorítmico produce determinados resultados. La opacidad algorítmica puede surgir por complejidad técnica, por decisiones de diseño o por restricciones contractuales impuestas por proveedores privados (Burrell, 2016). La transparencia no significa revelar el código completo, sino más asegurar que ciudadanos, supervisores y funcionarios puedan acceder a explicaciones sobre criterios, variables y razonamientos del modelo.

Los principales indicadores de transparencia son: , documentación del sistema, posibilidad de auditorías externas, trazabilidad de decisiones y divulgación de criterios generales. La regla normativa exige que toda decisión automatizada sea explicable, trazable y revisable, garantizando así a la ciudadanía el contar con vías efectivas de imanación de errores.

Finalmente, la legitimidad es el pilar más extenso ya que analiza si el uso de la IA mantiene, menoscaba o fortalece la autoridad la autoridad con base en la participación, control y rendición de cuentas. Si bien la IA puede incrementar la confianza cuando mejora servicios y coordina servicios de manera eficiente, también puede menoscabarla cuando deja de lado el juicio humano, concentra el poder de decisión en actores no electos o limita la capacidad de los ciudadanos para impugnar una decisión (Beckman et al., 2024; Yeung, 2018).

Entre los factores que afectan la legitimidad están el desplazamiento de procesos deliberativos, la excesiva delegación y la dependencia tecnológica. Por otra parte, los indicadores de legitimidad son los niveles de confianza de las personas, la claridad institucional sobre el uso de la IA y la existencia de mecanismos de apelación.

Su límite normativo es claro, la IA no puede tomar el lugar del núcleo de deliberación democrática. La regla operativa señala que los sistemas algorítmicos deben estar bajo la supervisión y aprobación del juicio humano y que las instituciones deben garantizar la transparencia, participación y control o auditoría externa.

Juntos estos cuatro ejes o pilares no como una herramienta neutral, sino como un instrumento que redistribuye riesgos, responsabilidades y capacidades.

La eficiencia sin equidad genera procesos rápidos pero injustos, la equidad sin transparencia puede producir sistemas bien intencionados pero imposibles de entender y la transparencia sin legitimidad genera sistemas claros pero antidemocráticos.

En conclusión, el marco teórico aquí propuesto integra estas dimensiones dentro de un análisis coherente que identifica tanto la evaluación de riesgos como la elaboración de defensas para una gobernanza pública responsable en contextos de digitalización acelerada.

Gobernanza algorítmica operativa

La gobernanza algorítmica en el sector público requiere un enfoque que trascienda el inventario de herramientas y se estructure como una secuencia operativa capaz de orientar decisiones institucionales antes, durante y después del despliegue de sistemas automatizados. Este enfoque reconoce que los riesgos asociados al uso de inteligencia artificial (IA) no emergen solo en la fase de implementación, sino desde la concepción del proyecto y se extienden a su operación cotidiana y a sus efectos acumulativos. Entender esta secuencia permite convertir principios éticos y normativos en procedimientos verificables con umbrales mínimos y condiciones de no implementación que no permitan lugar a dudas. La bibliografía reciente relacionada con la ética algorítmica y políticas públicas remarca que la gobernanza efectiva depende de esta estructura progresiva (Kaminski & Malgieri, 2021; Raji et al., 2020; Shneiderman, 2022).

En la fase antes de la implementación, las instituciones deben decidir si un sistema algorítmico es oportuno y justificado para el problema público que se pretende abordar. Esto involucra realizar la evaluación de impacto algorítmico que tenga en consideración los riesgos distributivos, sesgos potenciales, efectos sobre los derechos, impactos para las poblaciones vulnerables y potenciales externalidades institucionales.

Un análisis riguroso precisa considerar opciones no algorítmicas garantizando que la decisión de implementar la IA no responda a presiones tecnológicas o incentivos institucionales distorsionados.

Durante esta etapa deben establecerse umbrales mínimos que determinen cuando un sistema no debe ser implementados ya sea por ausencia de datos representativos, falta de mecanismos de supervisión

humana, definición pobre de criterios de decisión o incapacidad técnica para realizar auditorías externas. Igualmente, debe exigirse la documentación exhaustiva del sistema, incluidos objetivos, bases de datos utilizadas, supuestos del modelo y criterios de evaluación previstos (Mitchell et al., 2019).

La anterior etapa precisa que los contratos con los proveedores tengan cláusulas que garanticen la transferencia de datos, acceso a documentos técnicos, integración y depósitos de garantía sobre todo si se está hablando de sistemas de alto impacto (Zouridis et al., 2020).

Todos estos requisitos reducen el riesgo de dependencia económica y tecnológica, además de que garantizan que se seguirá operando aún si se cambia de proveedor.

En el transcurso de la implementación de la IA, la gobernanza se garantiza teniendo en funcionamiento mecanismos de control humano, seguimiento y auditorías. La supervisión humana en punto clave representa uno de los pilares normativos de la literatura contemporánea (Shneiderman, 2022). Se debe señalar que no basta con la simple presencia de un operador, el diseño del sistema debe permitir la intervención efectiva de la persona, con información suficiente para que comprenda el proceso de la toma de decisión del modelo, y debe contar también con la capacidad de cambiar, revertir o anular decisiones automatizadas.

La trazabilidad, por su parte, exige que todas las decisiones, incluidas las recomendadas por el algoritmo, queden registradas de manera que puedan ser analizadas posteriormente. Esto incluye variables procesadas, versiones del modelo, parámetros utilizados y contexto operativo. La literatura ha señalado que la ausencia de trazabilidad constituye uno de los principales factores que hacen imposible auditar los sistemas automatizados (Felzmann et al., 2020). Asimismo, la fase de implementación debe incluir auditorías externas habituales que evalúen el desempeño del modelo, la estabilidad y los efectos distributivos no anticipados. Estas auditorías deben ser realizadas por

entidades externas que no tengan conflictos de interés con el proveedor o con la agencia que implementa la tecnología (Raji et al., 2020)

También debe garantizarse la publicidad ya que los ciudadanos tienen el derecho de saber los criterios generales del sistema, su propósito, variables utilizadas y límites de su operación. La opacidad algorítmica, cuando no es técnicamente necesaria, erosiona tanto la transparencia como la legitimidad institucional (Burrell, 2016).

La fase posterior al despliegue se orienta a monitorear efectos acumulativos, corregir sesgos emergentes, evaluar impactos distributivos y garantizar mecanismos reales de apelación. Los sistemas automatizados no son estáticos: aprenden, se adaptan y, en ocasiones, se degradan con el tiempo. Es por esto por lo que la gobernanza debe considerar procesos de reentrenamiento supervisado y métodos para desactivar temporalmente cuando se manifiesten desviaciones considerables. Los impactos distributivos necesitan monitoreo fijo a través de indicadores de equidad y comparaciones entre diferentes grupos de la población, así, si se identifican efectos negativos que no pueden contrarrestarse de manera razonable, se debe activar un protocolo de suspensión del sistema.

La bibliografía también recalca la importancia de mecanismos fuertes de apelación ciudadana (Almada, 2021), que no solo implica la recepción de reclamos, sino que garantice que toda persona afectada pueda solicitar revisión humana, recibir explicaciones que pueda entender y posteriormente tener una solución efectiva a su reclamo. Además, esta etapa precisa también de evaluaciones constantes de legitimidad que analicen la percepción pública, los niveles de confianza institucional y la claridad en la comunicación gubernamental sobre el uso de la IA (Beckman et al., 2024).

En conjunto, la gobernanza algorítmica operativa se concibe como un proceso secuencial continuo que asegura que la IA no reemplace el juicio humano ni erosione principios democráticos fundamentales. Su fortaleza radica en transformar principios normativos: equidad, transparencia, participación y control democrático, en procedimientos verificables que las

instituciones puedan implementar independientemente de su nivel de sofisticación técnica. La IA tiene la capacidad para fortalecer la capacidad del Estado, pero únicamente si se implementa en un entorno que permita una continua supervisión, auditorías externas, trazabilidad completa y formas de apelación efectivas. De esta manera, lejos de ser un simple complemento técnico, la gobernanza algorítmica es la condición que determina si la IA aporta al bienestar social, o por el contrario incrementa las desigualdades y debilita la legitimidad democrática.

Casos internacionales

La evidencia internacional demuestra que la adopción de sistemas de inteligencia artificial (IA) en la administración pública produce resultados heterogéneos que dependen de los mecanismos institucionales de supervisión, la calidad de los datos y la capacidad estatal para corregir errores. Dos experiencias, una en el sector educativo y otra en la protección social, permiten observar cómo los beneficios esperados pueden verse condicionados por dilemas de equidad, transparencia y legitimidad democrática, y ofrecen lecciones operativas para orientar el uso responsable de sistemas algorítmicos.

En la educación, varios países han implementado plataformas educativas con base en la IA para personalizar los procesos de aprendizaje, identificar rezagos y prevenir deserción escolar. Un caso bastante analizado es el uso de sistemas de recomendación pedagógica en instituciones públicas de América Latina, en donde la IA se usa para ajustar contenidos, sugerir actividades diferenciadas para cada estudiante o grupo de estudiantes y monitorear el progreso de los alumnos en tiempo real (Ferreira de Lima & Veiga, 2025). El objetivo principal ha sido mejorar la eficiencia de la educación, reducir la carga laboral a los docentes y optimizar la asignación de recursos pedagógicos.

De esta manera, las primeras evaluaciones evidenciaron mejoras en el rendimiento académico de los estudiantes con acceso a infraestructura digital y en contextos institucionales donde la IA se utiliza como un complemento a

la labor del docente, más no como un sustituto. No obstante, los beneficios no se distribuyen de manera equitativa ya que las brechas de conectividad, falta de formación de los docentes y la limitada representatividad de algunos grupos en datos de entrenamiento han derivado en patrones de exclusión.

Los estudiantes de zonas rurales, con menor acceso a dispositivos digitales o conexión a internet reciben recomendaciones menos precisas o fuera de tiempo, lo que incrementa las desigualdades existentes (Capraro et al., 2024). Además el funcionamiento de estas herramientas puede llegar a ser poco comprensible para algunos docentes lo que explica por qué algunos estudiantes reciben ciertas intervenciones y otros no.

Esta falta de claridad afecta la confianza en las instituciones educativas y puede generar percepciones de arbitrariedad. En consecuencia, las conclusiones en el ámbito educativo son claras, la personalización algorítmica requiere datos representativos, supervisión docente activa, transparencia sobre criterios de recomendación y formas en las que corregir los diferentes procesos de manera que se evite que la IA incremente las brechas educativas existentes sobre todo en el contexto latinoamericano.

Una segunda área importante para el análisis de la implementación de la IA es la protección social. Varios países han adoptado sistemas automatizados para focalizar subsidios, unificar registros de beneficiarios o encontrar deficiencias en solicitudes de ayuda estatal. En Brasil el uso de sistemas algorítmicos para gestionar el Cadastro Único, una base de datos que contiene información de millones de hogares permitió mejorar la eficiencia operativa del programa Bolsa Familia, reduciendo tiempos administrativos y mejorando la coordinación entre agencias públicas (De la Briere & Lindert, 2005). La automatización permitió identificar rápidamente los hogares más vulnerables y actualizar la información para la asignación de transferencias monetarias mejorando así la eficiencia y la planificación del gasto público.

Empero, estos sistemas también han generado desafíos importantes. Por ejemplo, varias auditorías han encontrado casos de exclusión injustificada

producto de errores en la clasificación automatizada, así como dificultades para que los ciudadanos afectados impugnen estos inconvenientes.

En algunos municipios, los algoritmos operaban con datos antiguos o incompletos lo que generaba sesgos distributivos y afectaba de manera desproporcionada a la población rural personas mayores o demás población vulnerable. Las limitaciones para explicar también salieron a la luz ya que varios usuarios no entendían porque habían sido excluidos de beneficios, y en muchos casos, funcionarios tampoco podían reconstruir correctamente el razonamiento del modelo.

La experiencia brasileña resalta que la legitimidad democrática de estos sistemas no solo depende de la precisión técnica, sino también de la capacidad de los ciudadanos de entender , impugnar y corregir los fallos producto de las decisiones automatizadas.

En ambos escenarios, tanto en la educación como en la protección social la adopción de la IA en servicios públicos solo genera beneficios sostenibles cuando existe una estructura institucional que garantice la supervisión humana, auditoría externa, transparencia y métodos de corrección de errores.

La eficiencia obtenida no justifica por sí sola los riesgos que la automatización genera en términos de equidad y confianza pública. Tanto así que la evidencia internacional demuestra que los sistemas algorítmicos deben entenderse como herramientas complementarias a la intervención humana, cuyo trabajo depende del diseño institucional y del compromiso con una gobernanza democrática que garantice que toda esta innovación tecnológica se orienta hacia el bienestar público y no hacia la reproducción de desigualdades estructurales.

Discusión crítica

La integración de inteligencia artificial (IA) en la administración pública revela tensiones profundas entre los beneficios operativos que ofrece la automatización y los principios democráticos que estructuran la acción

estatal. Los ejes analíticos desarrollados: eficiencia, equidad, transparencia y legitimidad, no operan de manera aislada: se interrelacionan y, en ocasiones, entran en conflicto. La discusión crítica permite examinar estas interacciones y establecer criterios operativos para decidir cuándo un sistema algorítmico es compatible con los valores que sostienen las instituciones públicas.

En primer lugar, la eficiencia es el principal motor para la implementación de la IA. En los casos analizados queda en evidencia que la automatización mejora la velocidad de procesamiento de información, optimiza recursos e incrementa la capacidad estatal para coordinar intervenciones complicadas. No obstante, la eficiencia se torna confusa cuando no se evalúan los efectos distributivos ya que un sistema puede ser muy eficiente desde un enfoque administrativo, y a la vez generar resultados injustos y que afecten de manera desproporcionada a algunos grupos de la población. Es así como la bibliografía señala que los sistemas que utilizan la IA tienden a reforzar patrones de desigualdad cuando los datos de entrenamiento presentan sesgos históricos o cuando los mecanismos de supervisión son insuficientes (Barocas & Hardt, 2020; Mehrabi et al., 2022). En resumen, la eficiencia es admisible únicamente cuando se encuentra bajo la tutela de la equidad y no cuando la desplaza o la compromete.

En segundo lugar, la equidad se entiende como el eje que limita y orienta la utilización de la IA en procesos de alto impacto. Por ejemplo, la evidencia empírica internacional manifiesta que incluso los sistemas técnicamente bien diseñados pueden producir formas de exclusión que permanecen invisibles cuando se sustentan en bases de datos incompletas, desactualizadas o con poca representatividad.

En conclusión, la equidad requiere no solo medir los efectos distributivos, sino plantear umbrales mínimos de aceptabilidad ya que cuando un modelo genera diferencias no justificadas entre grupos, o cuando no es posible identificar las causas de dichas diferencias, deben activarse mecanismos de veto. Por lo tanto, esta regla de equidad mínima toma fuerza con la necesidad de monitoreo habitual ya que los modelos pueden desmejorarse creando aún más sesgos con el tiempo.

En tercer lugar esta la transparencia, que se traduce en la condición que permite evaluar tanto la eficiencia como la equidad ya que sin documentación técnica adecuada, trazabilidad operativa y auditorías externas, resulta imposible distinguir entre errores que son relativamente sencillos de corregir y fallos estructurales.

La discusión sobre la transparencia no se puede limitar a tener acceso al código que genera la información digital, más bien trata de la capacidad de comprender cómo funciona el sistema, cómo es el proceso de la toma de decisiones automatizadas y permitir la supervisión humana que está al tanto del funcionamiento de esta tecnología (Burrell, 2016; Shneiderman, 2022).

Los dos casos analizados demuestran que la incapacidad para explicar el funcionamiento de la IA se traduce en desconfianza por parte de la ciudadanía hacia las institucionales, y por lo tanto, la legitimidad de las decisiones es cuestionada, a pesar de que estén tomadas con base en modelos técnicos sólidos. Teniendo esto en consideración, se concluye que ningún sistema algorítmico que afecte derechos o asignaciones puede permanecer oculto, por lo tanto, la transparencia es un requisito indispensable para la corrección de errores y la rendición de cuentas.

Por último, la legitimidad democrática es el eje que une y evalúa a los demás. La IA puede fortalecer la legitimidad a través del mejoramiento de los servicios, reduciendo la arbitrariedad y ampliando la capacidad estatal. Empero, también puede debilitarla cuando reemplaza al juicio humano, limita la participación o crea una dependencia tecnológica que concentra el poder sobre todo en actores no electos (Beckman et al., 2024; Yeung, 2018). Por lo tanto, la legitimidad depende del balance entre los beneficios administrativos y los costos democráticos. Un sistema eficiente pero sin transparencia es ilegítimo y un sistema equitativo pero imposible de auditar igualmente lo es. Consecuentemente, la legitimidad requiere explicar y justificar la toma de decisiones, garantizar la apelación efectiva y asegurar que la IA complemente, más no reemplace a la deliberación pública.

La integración de los cuatro ejes permite identificar las condiciones bajo las cuales se puede implementar la IA de forma responsable y sostenida. Con base en ello se presenta una matriz que articula criterios operativos provenientes del marco teórico y de la evidencia empírica.

La matriz recomienda que la adopción de la IA es aceptable cuando los beneficios de eficiencia son verificables, medibles y constantes, no existen desigualdades entre los diferentes grupos de la población, existen fuertes mecanismos de control y evaluación y se garantiza la constante supervisión humana en cada fase de la formulación, implementación y evaluación de políticas públicas y toma de decisiones.

Por el contrario, no se debe implementar la IA si no hay datos suficientes o no son representativos, se afecta de manera negativa a la población vulnerable, los procesos no son comprensibles o no existe la posibilidad de apelación a la toma de decisiones.

En conclusión, en el sector público el valor de la IA depende de la estructura social que la acoge que de su capacidad técnica. En consecuencia, el principal desafío es integrar los beneficios de la implementación de la IA en el sector público con los valores de legitimidad del Estado.

Conclusiones y propuestas

Las transformaciones que introduce la inteligencia artificial (IA) en la administración pública no pueden comprenderse solo desde su dimensión tecnológica. Su aplicación en procesos estatales rediseña principios básicos de acción pública, modifica las relaciones entre las instituciones y los ciudadanos, y transforma los parámetros a través de los cuales evaluamos la justicia, la eficiencia y la legitimidad democrática.

Primeramente, el estudio arroja que aunque la eficiencia es un incentivo primordial para la implementación de la IA, esta no puede ser un criterio autónomo de decisión. La eficiencia administrativa es legítima solo

cuando va de la mano con la equidad distributiva y la transparencia institucional.

La automatización sin salvaguardas redistribuye los beneficios de forma injusta, más en sistemas donde los datos tienen desigualdades históricas o donde no hay suficiente supervisión humana. Es por esto por lo que la eficiencia debe regirse de manera que ningún beneficio operativo justifique la adopción de sistemas logarítmicos que generen diferencias injustificadas o reduzcan la capacidad de apelación o control ciudadano.

Por otro lado, los casos analizados muestran que incluso en sistemas técnicamente sólidos se pueden generar exclusiones si los datos son insuficientes o cuando no se garantizan los mecanismos de corrección. En consecuencia, se propone que todo sistema algorítmico debe pasar por evaluaciones habituales de impacto distributivo mediante métricas que permitan evidenciar desigualdades emergentes entre grupos. La ausencia de datos representativos, herramientas de monitoreo constante o canales de revisión provocan un posible veto para su implementación.

La transparencia es la condición que tanto la eficiencia como la equidad se puedan evaluar en la práctica ya que sin trazabilidad, documentación y explicabilidad, no se puede entender el razonamiento del modelo, corregir errores o evaluar su funcionamiento. Es así como la transparencia se convierte en un principio operativo y no solo ético ya que la capacidad de auditar, explicar y revisar las decisiones automatizadas es una categoría mínima para mantener la confianza de la población.

Por último, la legitimidad democrática une y evalúa los demás ejes o pilares. La IA fortalece la legitimidad cuando mejora la calidad del servicio, reduce arbitrariedades y aumenta la consistencia de las decisiones. Sin embargo, se puede debilitar si se reemplaza el debate humano, se concentra el poder en actores no electos o se pone trabas a la apelación de decisiones automatizadas, por esto la legitimidad necesita de instituciones que garanticen la participación, supervisión constante y transparencia.

El artículo señala que la IA es factible con la gobernanza democrática únicamente cuando funciona dentro de institución que procure la equidad, transparencia y constante supervisión humana. Esta estructura institucional se evidencia en procedimientos que se puedan verificar, evaluaciones de impacto logarítmico previas, auditorías externas, mecanismos fuertes de trazabilidad y canales abiertos para la apelación. Sin estos elementos, la eficiencia es un argumento insuficiente y en muchos casos, incluso peligroso.

Finalmente, este artículo también genera aportes conceptuales ya que ofrece un marco integrado que articula mecanismos causales a través de los que la IA transforma la acción estatal, así como criterios normativos que se pueden aplicar a diferentes contextos institucionales.

En resumen, la IA ofrece una valiosa oportunidad para mejorar la capacidad del estado, sin embargo, también tiene riesgos que no se pueden ignorar. Por ende, se tiene el desafío de diseñar una estructura institucional que garantice que su implementación estará regido por los principios democráticos y que su principal objetivo es aumentar el bienestar ciudadano.

El Estado tiene la misión de implementar esta tecnología sin dejar de lado los principios y valores que dan legitimidad a las instituciones públicas. Teniendo presente que cuando se adoptan protecciones claras, estándares verificables y procesos de constante revisión, la IA se convierte en un socio estratégico para la gobernanza democrática, pero cuando estos elementos están ausentes, se puede convertir en una serie de desigualdades y desconfianza institucional.

Referencias

- Almada, M. (2021). Human oversight in automated decision-making: Principles and challenges. *AI & Society*, 36(4), 1257–1270. <https://doi.org/10.1007/s00146-020-01077-3>
- Barocas, S., & Hardt, M. (2020). *Fairness in machine learning*. MIT Press.

- Beckman, R., Müller, J., & Hansen, T. (2024). Democratic accountability in algorithmic governance. *Governance*, 37(1), 55–74. <https://doi.org/10.1111/gove.12710>
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
- Capraro, C., Hernández, L., & Paredes, M. (2024). Algorithmic personalization in Latin American public education. *Education Policy Analysis Archives*, 32(12), 221–240. <https://doi.org/10.14507/epaa.32.2411>
- De la Briere, B., & Lindert, K. (2005). *Reforming Brazil’s Bolsa Família Program* (Social Protection Discussion Paper No. 0534). The World Bank. <https://openknowledge.worldbank.org/handle/10986/13687>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press.
- Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2020). Transparency you can trust: Transparency requirements for artificial intelligence. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720921115>
- Ferreira de Lima, G., & Veiga, A. (2025). AI-supported pedagogical systems in Latin America: Equity and performance implications. *Journal of Learning Analytics*, 12(1), 44–63. <https://doi.org/10.18608/jla.2025.8123>
- Kaminski, M. E., & Malgieri, G. (2021). Algorithmic impact assessments under the GDPR. *Computer Law & Security Review*, 41, 105532. <https://doi.org/10.1016/j.clsr.2021.105532>
- Kroll, J. (2020). Outlining traceability: A principle for accountability in machine learning systems. *Harvard Journal of Law & Technology*, 33(2), 403–454. <https://doi.org/10.1145/3442188.3445937>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35. <https://doi.org/10.1145/3457607>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. <https://doi.org/10.1145/3287560.3287596>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *arXiv*. <https://doi.org/10.48550/arXiv.2001.00973>
- Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.
- Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4), 505–523. <https://doi.org/10.1111/regg.12158>

Zouridis, S., van Eck, M., & Bovens, M. (2020). Automated decision-making in the public sector: Between efficiency and democratic legitimacy. *Information Policy*, 25(3), 281–297. <https://doi.org/10.3233/IP-200238>

Emociones y política: Estrategia teórico-metodológica para el análisis de la emocionalización del discurso político en el contexto posdemocrático

Emotions and politics: A theoretical and methodological strategy for analyzing the emotionalization of political discourse in the post-democratic context

Ekhiñe Graell Larreta

ekhiagraell@gmail.com

UAM-X

ORCID ID 0009-0003-6831-396X

Edith Flores Pérez

eflores@correo.xoc.uam.mx

UAM-X

ORCID ID 0000-0003-4545-1529

Resumen: El estudio de la política contemporánea requiere incorporar la dimensión emocional como parte constitutiva del poder. En el contexto posdemocrático (Crouch, 2004; Mouffe, 2023), marcado por la desafección ciudadana, la mediatización y la personalización del liderazgo, las emociones se han convertido en un eje estructural de la comunicación política. Lejos de ser manifestaciones irracionales, los afectos operan como mecanismos de adhesión, identificación o exclusión que moldean la cultura política y reorganizan los vínculos entre ciudadanía, instituciones y discurso.

Este trabajo propone una estrategia teórico-metodológica para analizar la emocionalización del discurso político, articulando cuatro enfoques

complementarios: el análisis crítico del discurso (Fairclough & Wodak, 2000; van Dijk, 2009), que permite situar las emociones en las relaciones de poder; el *framing* (Chihu, 2012; Goffman, 2006), que identifica los marcos afectivos de interpretación; la retórica (Aristóteles, 2002; Charaudeau, 2021), que analiza la interacción entre *ethos*, *pathos* y *logos* en la persuasión política; y la semiótica del discurso (Dagatti, 2017; Gutiérrez & Plantin, 2010), que examina las oposiciones simbólicas donde se anclan los relatos emocionales contemporáneos.

Desde el marco del giro afectivo (Ahmed, 2015; Macón, 2013), las emociones se conciben como prácticas sociales, discursivas y performativas que producen sentido, legitiman el poder y posibilitan la acción colectiva. La política actual se organiza mediante narrativas afectivas y no solo racionales: la indignación, la esperanza o el miedo orientan las adhesiones y delimitan antagonismos. En este sentido, las emociones funcionan como operadores de legitimidad y conflicto, en un espacio público dominado por la lógica mediática y personalización.

El modelo propuesto permite observar cómo las emociones se enuncian, se encuadran y se institucionalizan en el discurso político, revelando su papel en la construcción de identidades y en la disputa por el sentido. Superar la dicotomía razón-emoción implica comprender los afectos como vectores de poder y de acción política. Así, el estudio de la emocionalización ofrece una vía para entender cómo los discursos mediáticos y electorales configuran las sensibilidades colectivas y reestructuran las formas contemporáneas de participación democrática.

Palabras clave: Emocionalización de la política; posdemocracia; discurso político; análisis crítico del discurso; retórica; framing.

Abstract: Contemporary political analysis must integrate the emotional dimension as a constitutive part of power. In the post-democratic context (Crouch, 2004; Mouffe, 2023), marked by civic disaffection, mediatization, and leadership personalization, emotions have become structural elements of political communication. Far from being irrational manifestations, affects operate as mechanisms of adhesion, identification, and exclusion that reshape political culture and reorganize the relationship between citizens, institutions, and discourse.

This paper proposes a theoretical-methodological strategy for analyzing the emotionalization of political discourse, combining four complementary approaches:

critical discourse analysis (Fairclough & Wodak, 2000; van Dijk, 2009), which situates emotions within power relations; framing theory (Chihu, 2012; Goffman, 2006), which identifies affective interpretive frames; rhetoric (Aristotle, 2002; Charaudeau, 2021), which examines the interaction between *ethos*, *pathos*, and *logos* in political persuasion; and discourse semiotics (Dagatti, 2017; Gutiérrez & Plantin, 2010), which explores symbolic oppositions (people/elite, order/chaos, hope/fear) underlying contemporary emotional narratives.

Drawing on the affective turn (Ahmed, 2015; Macón, 2013), emotions are understood as performative social and discursive practices that produce meaning, legitimize power, and enable collective action. Current politics is organized through affective rather than merely rational narratives: indignation, hope, and fear guide adhesion and delineate antagonism. Emotions thus function as operators of legitimacy and conflict within a public sphere dominated by media logic and personalization of politics.

The proposed model observes how emotions are enunciated, framed, and institutionalized in political discourse, revealing their role in the construction of identities and struggles for meaning. Overcoming the reason-emotion dichotomy requires understanding affects as vectors of power and political agency. In this sense, analyzing emotionalization offers a critical pathway to understanding how media and electoral discourses shape collective sensibilities and redefine the conditions of democratic participation.

Keywords: Emotionalization of politics; post-democracy; political discourse; critical discourse analysis; rhetoric; framing.

Introducción

En las últimas décadas, las democracias representativas o liberales han experimentado una profunda transformación marcada por la creciente desafección ciudadana hacia las instituciones políticas. Esta crisis de representación, a la que se le ha denominado posdemocracia (Crouch, 2004) o postpolítica (Mouffe, 1999), se manifiesta en el vaciamiento ideológico de los partidos tradicionales, el distanciamiento entre la ciudadanía y las

instituciones democráticas, así como la creciente presencia de discursos polarizados y marcadamente emocionales.

Esta tendencia, intensificada por el papel de los medios de comunicación y las redes sociales ha generado un entorno en el que lo político -principalmente, la política formal o institucional apela de manera directa a lo emocional y coloca a la sensibilidad en un campo de batalla (Gómez & Velasco, 2024). El “homo politicus” parece haber sido reemplazado por un “homo emóticus” (Cantero, 2018), cuya actuación se rige por los afectos, las pasiones y emociones muchas veces articuladas desde narrativas populistas.

La emocionalización de la política se acompaña de importantes procesos de transformación social en las sociedades actuales: el auge de lo autobiográfico, el culto a la intimidad y el hedonismo bajo la dinámica del consumismo y el espectáculo (Arfuch, 2016), el capitalismo emocional con el imperativo de la felicidad y la concepción mercantil del amor (Cabanas & Illouz, 2019; Illouz, 2007, 2012, 2019), el consumismo online y la capitalización del deseo (Sibilia, 2024), por mencionar algunos ejemplos. En este contexto, resulta indiscutible que las emociones constituyen un componente fundamental en las dinámicas sociales y la política contemporánea.

En el marco de una investigación¹ más amplia que analiza la dimensión emocional de las campañas electorales a la presidencia de México durante 2024², el objetivo de este artículo es presentar una estrategia teórico-

¹ El proyecto de investigación posdoctoral “La mercantilización de las emociones y la personalización de la política. Análisis del discurso mediático de la campaña electoral a la Presidencia de México en 2024” se desarrolla con la asesoría de la Dra. Edith Flores Pérez la Universidad Autónoma Metropolitana, Unidad Xochimilco (UAM-X), en el marco del programa de “Estancias Posdoctorales por México” perteneciente a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti).

² Las elecciones presidenciales de México de 2024 se realizaron el 2 de junio y contaron con tres candidaturas principales: Claudia Sheinbaum Pardo, por la coalición del en ese entonces gobierno en turno Sigamos Haciendo Historia (Morena-PT-PVEM); Xóchitl Gálvez Ruiz, por la coalición opositora Fuerza y Corazón por México (PAN-PRI-PRD); y Jorge Álvarez Máynez, por Movimiento Ciudadano, que compitió sin alianza. Sheinbaum resultó electa con alrededor del 59.7 % de los votos, frente al 27.4 % de Gálvez y el 10.3 %

metodológica que orienta el análisis de la dimensión emocional del discurso político en un contexto posdemocrático. Partimos de una concepción de la política como un espacio de acción inseparable de lo social y del reconocimiento del peso de lo emocional en su configuración (Arias Maldonado, 2016). Se sostiene que las emociones, como categorías de análisis social, pueden aportar una comprensión más profunda sobre estos fenómenos y facilitar la elucidación de los mecanismos de adhesión, identificación, polarización y exclusión que estructuran el campo de lo político en nuestras sociedades. El estudio del discurso político debe indagar en las emociones que lo configuran, porque tienen un origen y un impacto profundamente social (Sierra, 2015).

En sintonía con los aportes teóricos del giro afectivo (Macón, 2013; Macón & Solana, 2015) en diálogo con la teoría política contemporánea (Arias Maldonado, 2016; Crouch, 2004; Mouffe, 2023; Nussbaum, 2014; Pasquino, 2016; Scherer, 2024) recuperamos algunos principios del análisis crítico del discurso, el *framing*, la retórica y la semiótica para diseñar una estrategia teórico-metodológica como ruta de acceso a la dimensión emocional del discurso político. El estudio de los efectos de las emociones en lo político requiere comprender de qué manera funcionan, cómo circulan y se transforman para moldear tanto los cuerpos individuales como colectivos (Ahmed, 2015). Como expresiones discursivas, las emociones no sólo intervienen en el contenido de los discursos, sino también en su forma, su estética y su performatividad. Las emociones constituyen modos de producción de sentido, a través de los cuales se comunican y refuerzan relaciones mediadas por estructuras de poder, jerarquías y órdenes socioculturales (López, 2024).

Si bien, en el campo de la filosofía, pensadores como Adam Smith, Thomas Hobbes o Baruch Spinoza hablaron de la cuestión de las pasiones y su rol en la política, es en las últimas décadas del siglo XX, que la visión de la política centrada en la razón es impugnada con fuerza desde diversas

de Máynez, con una participación ciudadana cercana al 61 %, lo que implicó la continuidad del proyecto político encabezado por Morena (*Voto y elecciones. Elecciones 2024*, s. f.)

perspectivas como la teoría feminista y otras corrientes críticas -filosóficas, antropológicas, sociológicas- bajo el marco del denominado giro afectivo, que afirma que las emociones y los afectos son un mecanismo intrínseco de la política y la vida pública (Macón, 2013).

De acuerdo con la descripción de Macón (2013) y Macón & Solana (2015), el giro afectivo ha permitido posicionar el estudio de los afectos y las emociones como componentes centrales de la vida social, reivindicar su papel en la vida pública y en las formas de aproximación al pasado. El giro apuesta por la disolución de los dualismos razón/emoción, cuerpo/mente, esfera pública/privada, acción/pasión. Crítica la concepción que otorga una valencia positiva o negativa a las emociones, así como resignifica el carácter pasivo o despolitizador asignado a aquellos afectos considerados “feos” o indeseables como el trauma (Cvetkovich, 2018) o la melancolía (Flatley, 2008). También pone en duda la conexión automática entre los afectos percibidos como positivos y la acción política, como ocurre con el orgullo. En este sentido, Ahmed (2015) subraya el hecho de que las emociones y los afectos no son en sí mismos responsables de la opresión o la emancipación, sino que su análisis depende siempre del contexto.

En este marco, las emociones son prácticas sociales y culturales, constitutivas de la materialidad del cuerpo, de su carácter performativo y su vínculo con la agencia. Más que una definición cerrada, este marco implica una comprensión relacional, contingente y performativa de las emociones, la cual pone en el centro los encuentros sociales y las articulaciones entre cuerpos y colectividades en detrimento de su concepción como propiedades discretas, fijas o psicológicas (Depetris-Chauvin, 2019).

Como construcciones culturales, históricas y sociales, las emociones y los afectos se encuentran en la base de las relaciones sociales y políticas, por lo cual son dimensiones capaces de dar cuenta del lazo social a la vez que son articuladoras de la experiencia (Macón, 2013). Las emociones participan de la normalización de las relaciones de poder al tiempo que tienen el potencial para desarticularlas (Ahmed, 2015; Macón & Solana, 2015).

Desde esta perspectiva, el análisis de las emociones más que buscar definiciones sobre lo que son, debe orientarse hacia la comprensión de sus efectos performativos, en el marco de los valores y los códigos normativos que configuran las acciones individuales y colectivas estructuradas por las relaciones de poder. El análisis de la dimensión emocional del discurso político resulta fundamental para entender la cultura política de nuestra época (Gantús, Rodríguez & Salmerón, 2021). Lo cual implica a su vez, la reconfiguración del concepto de “ciudadano” y de “lo político” en diálogo constante con las emociones y su papel en la esfera pública.

Desde estos presupuestos teóricos, para alcanzar el objetivo propuesto, el artículo se organiza en tres secciones. En primer lugar, se describen algunos aspectos de la emocionalización de la política en contextos posdemocráticos y el papel de los medios y las redes sociales en la construcción de discursos emocionales. En segundo lugar, se describe la estrategia metodológica diseñada en el marco de la investigación posdoctoral en curso para el análisis de la dimensión emocional del discurso político en el contexto posdemocrático, y en tercer lugar, el artículo cierra con un apartado de reflexiones finales sobre la relevancia de esta perspectiva teórico-metodológica para el estudio de las emociones en las democracias contemporáneas.

La emocionalización de la política en el contexto posdemocrático y el papel de las nuevas tecnologías de información

En las últimas décadas, la democracia occidental ha experimentado profundas transformaciones, sobre todo después de la crisis financiera de 2008. Este nuevo contexto, denominado postpolítica o posdemocracia, es reflejo del debilitamiento de los fundamentos esenciales de la democracia como la igualdad y la soberanía popular (Mouffe, 2023), así como la libertad, lo que ha erosionado los principios centrales del pensamiento ilustrado (Guerrero, 2019). Esta crisis de la esfera democrática ha dado lugar a un escenario en el que tanto la izquierda como la derecha coinciden en aceptar que no existe alternativa alguna al actual modelo neoliberal, generando una ausencia de antagonismo ideológico real en el plano político. Como resultado, las

campañas electorales actuales se limitan a la simulación de diferencias políticas, sin proyectos auténticamente diferenciados que realmente signifiquen distintas opciones políticas para el electorado (Mouffe, 2023; Žižek, 2010).

Colin Crouch (2004) denomina posdemocracia a este contexto en el que si bien, las prácticas democráticas persisten, su esencia ha sido cooptada por pequeñas élites políticas y económicas. Dichas élites son las que concentran el poder y las que reducen la participación efectiva de la ciudadanía. De esta manera, la globalización, los cambios en la estructura de clases y la preeminente disminución de la influencia religiosa han transformado las formas tradicionales de conexión entre políticos y ciudadanos. Mientras que en las democracias tradicionales la base de la política era la identificación social, en la posdemocracia los nuevos ejes de la participación son el *marketing* político y la imagen personal (Crouch, 2020). Esto ha desplazado la ideología política de las campañas electorales, sustituyéndola por estrategias de comunicación masiva.

Josep M. Colomer (2023) señala que la globalización ha debilitado a los Estados-nación, lo que ha deteriorado la gestión pública. Como consecuencia, el electorado de las democracias occidentales busca alternativas fuera de los partidos tradicionales. Nancy Fraser (2023) también coincide en diagnosticar la crisis de la democracia, pero atribuye sus causas al sistema capitalista en su totalidad. A esto se añade la creciente relevancia de los nuevos medios de comunicación en la política contemporánea, los cuales han alterado radicalmente la relación entre lo público y lo privado. En esta nueva configuración, la ideología tiene mayor influencia en las prácticas cotidianas que en la esfera política formal (Žižek, 2024). Esto ha generado un auge de liderazgos políticos que son “más expresivos que ideológicos” (Gallardo, 2017, p. 191), desplazando la ideología política a un segundo plano y priorizando la imagen y el mensaje como herramientas centrales del *marketing* político.

En este contexto, la política se ve crecientemente mediada por estrategias de emocionalización: discursos, imágenes y narrativas diseñadas

para provocar reacciones emocionales inmediatas, que a su vez orientan conductas electorales y adhesiones ideológicas. Mouffe (1999, 2023) recuerda que toda identidad política se articula mediante pasiones colectivas, y que la tarea de la teoría democrática no es negar esta dimensión, sino encauzarla en marcos agonísticos donde el conflicto no se suprima, sino que se gestione de forma productiva para la pluralidad. Sin embargo, en las dinámicas posdemocráticas, la apelación afectiva tiende a canalizarse hacia la polarización excluyente y la personalización extrema del poder.

Otra de las características del contexto posdemocrático es la polarización “un fenómeno multifacético que no solamente significa una división de posturas respecto a temas de interés y relevancia política, sino, de manera más específica, al distanciamiento de éstas” (Moreno, 2024, p. 195), de manera extrema o únicamente como un distanciamiento entre los grupos políticos. La polarización es un fenómeno político actual que está estrechamente relacionado con la circulación de emociones en la esfera pública a través de discursos políticos afectivos: “La polarización política es, en esencia, una polarización afectiva. [...] En contextos de polarización, la diferencia ideológica entraña una intensa implicación emotiva por la cual las adhesiones o rechazos políticos comienzan a impregnar la interacción interpersonal de los ciudadanos” (Ramírez & Falak, 2023, p. 370).

A esto se suma el fenómeno denominado como la personalización de la política que se da cuando los atributos personales de un político son considerados más importantes que su biografía política, su experiencia en política, su desempeño, sus programas y sus promesas, sea por él o por quienes se fijan en él (Pasquino, 2016). La personalización del poder se encuentra estrechamente relacionada con la mediatización de la política, que ha traído consigo que los líderes políticos actuales actúen más “como celebridades que ofrecen información y representaciones más ligadas al entretenimiento, es decir, al *infoteinment*, para intentar ofrecer una nueva cara de las figuras políticas” (Siria, 2019, p. 99) lo que ha provocado el nacimiento de la denominada política pop. De acuerdo con la propuesta de Van Aelst, Sheafer y Stanyer (2012), la personalización engloba cuatro elementos

fundamentales: la cobertura de la familia, la crianza y el pasado, el tiempo libre (los pasatiempos, las vacaciones) y la vida amorosa.

Uno de los aspectos más relevantes de esta era postpolítica es la cada vez más creciente dependencia del *marketing* político y la comunicación en la esfera pública. De acuerdo con Chantal Mouffe (2023), las características de este escenario incluyen lo siguiente: 1) Bipartidismo: las únicas alternativas políticas viables son de centroizquierda o centroderecha, por lo que se excluye cualquier otra opción fuera de ese espectro. 2) La precarización de la clase media, que dinamita el principio de igualdad, uno de los pilares fundamentales de la democracia representativa. 3) El dominio del sector financiero sobre el productivo. De acuerdo con la autora, en la postpolítica, cualquier idea política hallada fuera del centro es considerada como populista.

Chantal Mouffe (1999), considera a las emociones como un elemento político importante y potencialmente democrático. Teniendo como base las pasiones colectivas por parte de partidos populistas de derecha o movimientos fundamentalistas y terroristas que se movilizan en un modo antagónico y moralista de “lo correcto y lo incorrecto”, creando una distinción entre “nosotros” y “ellos” sostiene que la teoría democrática y la política, al ignorar las emociones, tiene como resultado el auge de los populismos de derecha. Se opone al concepto democrático de un consenso racional, favorecido por las teorías deliberativas, y sostiene en cambio que la política y lo político no pueden entenderse sin reconocer las pasiones como fuerzas impulsoras. De esta manera, las emociones son una herramienta importante en la lucha contra la apatía o la alienación y, por lo tanto, un potencial impulsor de la participación política.

Así, a partir de la introducción del *marketing* en la comunicación, incluyendo la comunicación política, el pragmatismo ha permeado los mensajes y la imagen de los políticos. Comenzó la concepción de la democracia como un mercado político en el que los votantes podían escoger libremente a sus representantes, diferenciados ya, no por su ideología política, sino por la forma en la que mercadológicamente fue construida su imagen. En esta era posdemocrática, lo importante es la forma y no el fondo: al

político contemporáneo se le reconoce “por sus ocurrencias, por su eslogan de campaña, por su avatar, por su atuendo, por sus actitudes, por sus tuits y *posts* en las redes digitales, pero no -ni falta hace- por sus proyectos y sus ideas” (Guerrero, 2019, p. 38). 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, el discurso ideológico ha quedado relegado de las campañas y los discursos políticos, aunque la ideología política continúa en forma de emociones como el odio (Ahmed, 2015). Como lo planteó Ramón Maíz (2010) “sin emociones no hay empatía posible y sin empatía no hay en rigor seres sociales ni ideologías compartidas, ni identidades colectivas, ni eventual acción de consumo. Sin emociones no existe indignación ni compasión ante la crueldad” (p. 42).

En conjunto, los aportes revisados ofrecen una serie de categorías analíticas para observar y explicar el fenómeno de la emocionalización de la política en el contexto posdemocrático, que podemos clasificar en tres bloques. En primer lugar, los trabajos enmarcados en el giro afectivo (Ahmed, 2015; Depetris-Chauvin, 2019; López, 2024; Macón, 2013; Macón & Solana, 2015; Nussbaum, 2014) muestran que emociones como el miedo, la ira/indignación, la esperanza, el orgullo, la compasión o la decepción ejercen una importante función en el vínculo entre la ciudadanía y el poder, pero que sus posibles efectos políticos dependen del contexto en el que se producen. De igual manera, en ese mismo sentido, plantean que no existen unas emociones mejores que otras para la movilización y la acción política. Las emociones, además, se encuentran atravesadas por la lógica del capitalismo (Cabanas & Illouz, 2019; Illouz, 2007, 2012, 2019; Sibilia, 2024) y la centralidad de la dimensión afectiva en los fenómenos sociales (Arfuch, 2016; López, 2024), lo que obliga a analizarlos como productos históricos, socialmente producidos y ligados a prácticas, discursos y tecnologías.

En segundo lugar, la literatura sobre posdemocracia y postpolítica (Colomer, 2023; Crouch, 2004, 2020; Fraser, 2023; Guerrero, 2019; Mouffe,

1999, 2023; Žižek, 2010, 2024) muestra que el escenario político posdemocrático se organiza a través de una lógica afectiva, donde la polarización, la personalización del liderazgo y la mediatización intensifican la circulación de emociones. La crisis de representación, derivada de las consecuencias del modelo neoliberal y la erosión de la soberanía popular se traducen en discursos que: (a) problematizan la realidad en términos de crisis, amenaza o decadencia; (b) atribuyen causalidades a actores concretos (élites, enemigos internos, adversarios ideológicos); (c) proyectan futuros de salvación, restauración o colapso; y (d) organizan identidades colectivas a partir de polarizaciones afectivas entre “nosotros” y “ellos” (Fraser, 2023; Mouffe, 1999, 2023; Žižek, 2010, 2024). Los estudios sobre polarización afectiva (Moreno, 2024; Ramírez & Falak, 2023; Scherer, 2024) muestran que estas dimensiones discursivas no solo separan posiciones ideológicas, sino que intensifican distancias emocionales entre grupos, convirtiendo las diferencias políticas en desacuerdos morales y afectivos.

En tercer lugar, los análisis sobre mediatización, personalización y *marketing* político (Aira, 2020; Cantero, 2018; Gallardo, 2017; Gómez & Velasco, 2024; Guerrero, 2019; Pasquino, 2016; Siria, 2019; Van Aelst, Sheaffer & Stanyer, 2012) permiten precisar las funciones políticas de las emociones en el contexto posdemocrático. La política pop, el *infoteinment* y la construcción de liderazgos “más expresivos que ideológicos” (Gallardo, 2017) muestran que las emociones cumplen, al menos, cuatro funciones: (a) movilizar o desmovilizar a la ciudadanía; (b) legitimar proyectos, gobiernos o candidaturas mediante afectos positivos (Arias Maldonado, 2016; Nussbaum, 2014); (c) deslegitimar a adversarios mediante emociones negativas (Ahmed, 2015; Maíz, 2010); y (d) simplificar y hacer consumible la complejidad política a través de mensajes breves, personalizados y altamente emocionalizados, ajustados a la lógica de las plataformas digitales (Aira, 2020; Guerrero, 2019; Sibilia, 2024). En este sentido, los medios de comunicación masiva y sobre todo, las redes sociales, han intensificado la circulación y la visibilidad de estas funciones emocionales, al privilegiar contenidos que producen impacto afectivo inmediato.

Hacia una propuesta metodológica para analizar la emocionalidad del discurso político en la política contemporánea

La revisión teórica permite sostener que las transformaciones contemporáneas de la política -marcadas por la mediatización, la personalización, la polarización y la emocionalización- se expresan de manera privilegiada en el discurso. Se trata de un escenario en el que el discurso político no solo transmite información, sino que organiza percepciones, genera adhesiones políticas y legitima posiciones ideológicas. Por esta razón, el análisis de la política actual debe analizarse de manera integral, a través de las condiciones de producción, los marcos de interpretación que activa, las estrategias de persuasión que se utilizan y las oposiciones afectivas que construye. La propuesta metodológica que se presenta a continuación se fundamenta en la articulación del análisis crítico del discurso, el *framing*, la retórica y la semiótica que permiten observar el discurso político contemporáneo de manera situada, para analizar las dinámicas de acción política en el contexto posdemocrático.

El análisis crítico del discurso es una metodología que examina el lenguaje en tanto práctica social, enmarcada en relaciones de poder y en contextos políticos, sociales y culturales concretos (Fairclough & Wodak, 2000; van Dijk, 2009) y puede resultar útil para el análisis del discurso político emocional. El discurso político no solo transmite información: moviliza emociones, legitima posiciones ideológicas y configura imaginarios colectivos. Para el análisis de la emocionalidad en la política contemporánea en particular, interesa analizar cómo se construyen, circulan y utilizan las emociones. De acuerdo con van Dijk (2009), el poder que se reproduce mediante los discursos es principalmente, persuasivo y “quienes controlan el discurso, pueden controlar indirectamente las mentes de la gente” (van Dijk, 2009, p. 30), determinando qué conocimientos adquieren, qué opiniones generan, así como sus actitudes y la ideología que detentan, pero, además, qué emociones provocan o movilizan.

La palabra posee el poder de hacer referencia a todo un campo semántico, es decir, a todo un conjunto de significados. “Los nombres que

damos a las cosas importan, a pesar de que [...] un nombre es, en principio, arbitrario. Nombrar es enmarcar” (Schröder, 2014, p. 173). El ACD [análisis crítico del discurso] considera que toda realidad es, en sí misma, un sistema comunicativo, entendiendo como comunicación a todo aquel sistema que posee una forma codificada mediante la cual se transmiten ideas, actitudes, comportamientos y/o mensajes (van Dijk & Rodrigo, 1999). Por tanto, nos permite comprender los contenidos de un determinado mensaje, en la que tiene más importancia conocer los alcances de los mensajes que la comunicación en sí misma, como es el caso de un discurso político (Fairclough & Wodak, 2000).

Concebir al discurso como una práctica social, tal y como lo plantea el ACD, implica reconocer que existe una relación recíproca entre lo que se dice y el mundo social en el que se produce. El discurso se encuentra moldeado por las instituciones, los contextos y las estructuras sociales y, al mismo tiempo, lo que se dice y la forma en la que se dice también tiene la capacidad de modelar a la sociedad, influyendo en la manera en la que entendemos nuestra realidad, cómo nos relacionamos entre nosotros y las identidades colectivas que se construyen. Cuando trasladamos este planteamiento al estudio del discurso político actual, cargado de afectividad, esa lógica se hace evidente. Las emociones situadas en los discursos políticos son una respuesta a condiciones sociales concretas, ya sea la desconfianza hacia las instituciones o el malestar económico. Sin embargo, no sólo se limitan a reflejar esas condiciones, sino que las reconfiguran. Al apelar a las emociones, el discurso político orienta cómo interpreta la ciudadanía los problemas, identifica a los responsables y se posiciona frente a ellos.

En términos metodológicos, esto significa que el ACD nos puede permitir analizar la función de las emociones como mediadoras entre las estructuras sociales y la acción política. Tomando en cuenta los planteamientos de esta perspectiva, sería lo social lo que determinaría las emociones que aparecen en el discurso, como, por ejemplo, el miedo frente a la inseguridad o la esperanza ante un proyecto de cambio. De igual manera, es el discurso emocional el que organiza y transforma lo social, creando

identidades políticas como “nosotros” frente a “ellos”, legitimando ciertas visiones de futuro y excluyendo a otras. En suma, el ACD aplicado a la emocionalidad del discurso político permite revelar que las emociones no son simples expresiones subjetivas, sino herramientas de construcción social y política que pueden contribuir tanto a la reproducción del orden existente (como puede ser el reforzamiento de prejuicios o jerarquías) como a su transformación, al movilizar nuevas maneras de pensar y de actuar colectivo.

En este punto conviene precisar que la emoción no opera como una simple decoración del argumento de un discurso político, sino como un recurso movilizador (Ahmed, 2015). Siguiendo a Christian Plantin (2014), la argumentación emocional concibe las emociones como formas de prueba socialmente aceptadas, capaces de orientar inferencias y anclar acuerdos o desacuerdos sin necesidad de apelar siempre a evidencias factuales. Las emociones actúan como marcadores de relevancia, señalando qué importa y por qué; como operadores evaluativos, jerarquizando valores en juego; así como dispositivos de posicionamiento, definiendo cercanías y distancias entre actores. Esta perspectiva desplaza la dicotomía simplificadora de razón-emoción, obligando a describir cómo se construyen las “buenas razones afectivas” en una escena pública dada (Plantin, 2014). En un discurso de campaña, por ejemplo, la indignación no solo intensifica el tono: funciona como prueba de injusticia preexistente, habilitando derechos a exigir y obligaciones a reparar; la esperanza, por su parte, opera como garantía anticipada de plausibilidad de un futuro mejor, encuadrando como razonables acciones que, sin dicho soporte afectivo, parecerían arriesgadas o costosas.

Ahora bien, si aceptamos que las emociones pueden fundar y no únicamente acompañar argumentos, entonces el análisis metodológico debe identificar escenas argumentativas: momentos, formatos y géneros discursivos donde se negocian definiciones de la situación (crisis/normalidad, amenaza/oportunidad, élite/pueblo) y se indexan emociones a categorías políticas (pueblo honesto/élite corrupta; víctimas/victimarios; cuidadores/depredadores). En estas escenas, las huellas lingüísticas — adjetivaciones intensificadoras, metáforas corporales y bélicas, oposiciones morales, deícticos de proximidad (“aquí”, “nosotros”), marcadores de certeza

(“sin duda”, “está claro”)— materializan la pretensión de validez afectiva del discurso y, en consecuencia, deben rastrearse sistemáticamente (Plantín, 2014). No debemos olvidarnos de que, “para la teoría retórica, es imposible estudiar la argumentación desatendiendo las emociones que están ligadas a las situaciones argumentativas de base, el debate político y la confrontación judicial” (Plantín, 2012, p. 94).

De acuerdo con Silvia Gutiérrez y Christian Plantín (2010) ciertas emociones deben entenderse como movimientos enunciativos que presuponen un destinatario, distribuyen papeles (quién puede demandar, quién debe responder, quién queda interpelado) y modulan contratos de lectura (qué se considera aceptable o razonable en ese intercambio): “El objetivo es precisar los principios (o *topoi*) que aseguran la coherencia del discurso que conmueve” (Gutiérrez & Plantín, 2010, p. 50). En términos metodológicos, esto supone observar la orientación de la acción que se relaciona con lo declarado: cuando un liderazgo alude al dolor, ¿a quién habilita como agente reparador?; cuando se indigna, ¿a quién atribuye causalidades y qué reparación reclama?; cuando promete esperanza, ¿qué futuro de credibilidad y de costos ofrece? Esto obliga a codificar no sólo “qué emoción aparece” en el discurso político, sino qué hace esa emoción: ¿convoca?, ¿amenaza?, ¿consolida un “nosotros”?; ¿desacredita al adversario?

Por su parte, Mario Dagatti (2017) ha mostrado cómo la política contemporánea se sostiene en gran medida en la exposición emocional de los liderazgos. En sus estudios sobre discurso político y medios, explica que las emociones no solo acompañan el mensaje, sino que forman parte de la construcción pública del liderazgo. La manera en que un político se muestra—su tono, gestos, pausas, silencios o cercanía con la audiencia— incide en cómo se percibe su autoridad y su credibilidad (Dagatti, 2017). Desde esta perspectiva, analizar el discurso político implica también observar cómo los medios y las plataformas transforman esa relación afectiva: un mitin televisado, una entrevista, una publicación en redes o un video corto no despiertan las mismas emociones ni permiten las mismas formas de vínculo. Cada formato impone condiciones distintas de visibilidad y respuesta, que

influyen en la intensidad, el tipo y la dirección de los afectos que circulan entre líder y audiencia. Por eso, en el plano metodológico, resulta necesario atender no sólo al contenido verbal, sino también al modo en que se comunica: los encuadres, repeticiones, silencios, énfasis o primeros planos. Estos recursos pueden reforzar o suavizar la carga emocional del discurso y consolidar ciertos estilos de liderazgo —por ejemplo, el tono protector, combativo o conciliador— que se vuelven reconocibles para el electorado (Dagatti, 2017).

Por ende, otro aspecto que considerar para el análisis del discurso emocional es el *framing*, que se da “cuando, en el curso de la descripción de un evento o un problema, el narrador enfatiza una serie de consideraciones que origina que la audiencia se enfoque en esas consideraciones al construir sus opiniones” (Chihu, 2012, p. 88). No hay que olvidar que la comunicación es un acto bilateral entre los productores y los receptores, y que no es posible entender el proceso comunicativo sin el contexto en el que se genera. Así, cuando se produce un mensaje, además de considerar lo que los emisores desean transmitir, toman en cuenta cómo podrían entender los receptores dicho mensaje, puesto que “la forma simbólica del mensaje ocupa un lugar privilegiado dentro del intercambio comunicativo; y [...] los momentos de ‘codificación’ y ‘descodificación’, aunque ‘relativamente autónomos’ en relación con el proceso en su totalidad, son instantes determinantes” (Hall, 2004, p. 218).

La producción del mensaje está, en buena medida, influenciada o fijada por el propio sistema político, social y cultural en el que se encuentra enmarcada. De esta manera, la producción de la noticia y la recepción de ésta están estrechamente relacionadas y, aunque nadie pueda garantizar que todos los receptores entenderán el mensaje de manera homogénea, los productores tienden a generar significados dominantes o preferentes que es posible analizar. Esos significados dependen de las clasificaciones que establecen los propios miembros de una sociedad sobre su entorno político, social y cultural que se conforman en una disposición cultural dominante, englobando los distintos campos de la vida social en categorías connotativas que poseen significados preponderantes o dominantes (Hall, 2004). Esos significados

generan patrones de lectura predominantes en un orden social determinado y, por tanto, generan encuadres o *frames* influyentes.

Una de las referencias esenciales en torno al discurso político se encuentra en *Retórica* de Aristóteles. En esta obra clásica, el filósofo griego estableció los tres modos de persuasión del discurso político: el *ethos*, basado en el carácter del orador, hace referencia a la imagen de credibilidad que el orador proyecta de sí mismo, se trata de la forma en la que el líder se presenta como digno de confianza, competente y moralmente aceptable; el *pathos*, que apela a las emociones del auditorio, busca conmover, movilizar afectos y generar identificación, es el eje de la movilización emocional que sostiene las adhesiones políticas; y el *logos*, que se realiza a través del razonamiento y la lógica, se relaciona con la dimensión racional y argumentativa del discurso, son las pruebas, razones y evidencias que se exponen para sostener una postura o una decisión política. De acuerdo con sus propias palabras, “de los argumentos procurados por el discurso hay tres especies: unos residen en el comportamiento del que habla; otros, en poner al oyente en una determinada disposición; otros, en el propio discurso, por lo que demuestra o parece demostrar” (Aristóteles, 2002, p. 53). En términos actuales, los primeros se refieren al emisor, quien debe dar una imagen de sí que genere confianza, esto es *auctoritas*, al público; los segundos al receptor, a quien se debe predisponer de cierta manera, provocando en él determinados estados de ánimo; y los terceros al mensaje, el contenido.

Patrick Charaudeau (2021) plantea que debido al predominio de las emociones en la política contemporánea, donde las imágenes y las emociones pesan más que la razón y los valores, hay una preeminencia del *ethos* y *pathos* en el discurso político, quedando relegado el *logos*: “sin dejar de ser una mezcla de estos tres componentes, se desplazó progresivamente desde el lugar del *logos* hacia el del *ethos* y el *pathos*, desde el lugar del contenido de los argumentos hacia el de su puesta en escena” (p. 53), asemejándose al discurso publicitario y mediático. De acuerdo con el autor, el discurso político está compuesto por la instancia política, la instancia adversaria, la instancia ciudadana y la instancia mediática. La primera, es la que produce el discurso con intención de

persuadir, se encarna en actores como partidos, líderes e instituciones de gobierno. Tiene como principal objetivo conquistar legitimidad, apoyo y credibilidad. La segunda, está formada por los contrincantes políticos y tiene la función de refutar, cuestionar y desacreditar los argumentos y la imagen (el *ethos*) de la instancia política, asegurando el conflicto y la confrontación de visiones en el juego democrático. La tercera, corresponde al público receptor, es decir, al electorado y a la ciudadanía, que es quien otorga o retira legitimidad, evalúa el *ethos* del político, responde emocionalmente (*pathos*) y examina los argumentos (*logos*). La última actúa como mediadora entre la política y la ciudadanía, seleccionando, reformulando, jerarquizando y haciendo espectaculares los mensajes, de manera que no sólo transmite, sino también condiciona el sentido mediante encuadres, imágenes y narrativas.

A partir del diálogo entre el análisis crítico del discurso, el *framing*, la retórica y la semiótica del discurso, se propone una estrategia metodológica de cuatro pasos para analizar la emocionalización del discurso político. Cada herramienta aporta una dimensión complementaria: el ACD examina el poder y las estructuras sociales que enmarcan el discurso; el *framing* permite identificar los marcos de interpretación; la retórica aborda los mecanismos de persuasión; y la semiótica explica cómo los relatos políticos construyen sentido mediante oposiciones simbólicas y narrativas.

1. ACD: contexto, actores y poder. Consiste en situar el discurso dentro de su contexto social, político y mediático (Fairclough & Wodak, 2000; van Dijk, 2009). Aquí se identifican: Quién habla, ante quién y desde qué posición de poder o legitimidad. Las condiciones de producción y circulación del discurso. Los temas dominantes y las representaciones sociales que el discurso reproduce o desafía. Este paso permite comprender qué relaciones de poder y dominación se inscriben en el discurso emocional y cómo el lenguaje contribuye a reforzarlas o cuestionarlas.

2. *Framing*: construcción de problemas, causas y soluciones. Se basa en la noción de *framing* (Chihu, 2012, 2024; Goffman, 2006), que permite identificar cómo el discurso selecciona ciertos elementos de la realidad y los jerarquiza emocionalmente. Para cada fragmento o unidad discursiva se

analizan tres aspectos: 1) *Frame* de problema: ¿Qué situación se presenta como urgente o amenazante?, 2) *Frame* de causalidad: ¿A quién o a qué se atribuye la responsabilidad?, 3) *Frame* de solución o futuro: ¿Qué acción o visión se propone como salida? Cada uno de estos marcos activa emociones específicas y orienta la acción política de los actores involucrados. Este paso revela la estructura interpretativa del discurso y su anclaje afectivo.

3. Retórica: *ethos*, *pathos* y *logos* como estrategias de persuasión emocional. Consiste en examinar los recursos retóricos del discurso, especialmente la interacción entre *ethos*, *pathos* y *logos* (Aristóteles, 2002; Charaudeau, 2021; Plantin, 2014). *Ethos*: cómo se construye la imagen del hablante. *Pathos*: qué emociones se evocan en el público y con qué propósito. *Logos*: qué argumentos factuales acompañan o justifican esas emociones. En esta fase se observan las marcas lingüísticas y discursivas (metáforas, intensificadores, repeticiones, tono, silencios) que convierten la emoción en argumento.

4. Semiótica del discurso: oposiciones y relatos emocionales. Profundiza en los relatos y oposiciones simbólicas que sostienen la carga afectiva del discurso, desde una mirada semiótica (Dagatti, 2017; Gutiérrez & Plantin, 2010). Aquí se identifican: Las oposiciones narrativas centrales (pueblo/élite, orden/caos, pasado/futuro, nosotros/ellos). Los relatos dominantes (crisis, salvación, agravio, renacimiento, comunidad). Los valores y emociones asociadas (miedo, esperanza, indignación, orgullo). Estos relatos organizan el discurso como una trama emocional que orienta la interpretación política y legitima la acción. La emoción se entiende, entonces, no sólo como un estado, sino como una estructura de sentido que da coherencia narrativa al proyecto político.

Esta integración metodológica permite analizar la emocionalización del discurso político en varios niveles: desde las estructuras de poder que lo producen, hasta los relatos simbólicos que lo legitiman. Al combinar ACD, *framing*, retórica y semiótica, se obtiene una mirada multinivel, capaz de explicar cómo las emociones se formulan lingüísticamente, se enmarcan narrativamente, se argumentan retóricamente y se fijan simbólicamente en la

cultura política. Esta propuesta facilita analizar no sólo qué dicen los actores políticos, sino cómo lo dicen, por qué lo dicen y qué emociones movilizan y para qué. El uso de estas herramientas metodológicas puede ayudar a analizar la lógica de la política contemporánea, en la que en el discurso no prevalece la transmisión de información o argumentos factuales, sino que busca, eminentemente, movilizar emociones y generar adhesiones políticas.

Conclusiones

Como se planteó a lo largo del texto, sostenemos que comprender la política contemporánea exige incorporar de manera sistemática la dimensión emocional como parte constitutiva del poder y no como un residuo “irracional” del proceso democrático. En este contexto posdemocrático donde la desafección y la mediatización reorganizan los vínculos entre ciudadanía, élites e instituciones, las emociones operan como dispositivos que orientan la atención, estructuran juicios y valores y pueden activar o neutralizar la acción política. De ahí que la política formal actual, lejos de agotarse en argumentos políticos que giran en torno a un programa o proyecto de gobierno, es decir, la discusión de propuestas políticas concretas se despliega en dinámicas de circulación y gestión de emociones que sostienen adhesiones, antagonismos e identidades colectivas.

A partir de esta premisa, se desarrolló un marco teórico que analiza la relación entre emociones y política, mostrando que cada emoción moviliza acciones y percepciones políticas determinadas. La ira, el miedo, la tristeza, el orgullo, el entusiasmo, el amor y la decepción actúan de manera diferenciada: algunas favorecen la movilización y la identificación colectiva, mientras que otras generan desconfianza, apatía o repliegue ciudadano siempre en el marco del contexto en el que se producen. Estas dinámicas se fortalecen con la comunicación digital y la lógica de las redes sociales, donde la exposición constante y la búsqueda de visibilidad intensifican la personalización y la polarización emocional. No obstante, esto no implica la desaparición de la capacidad ciudadana de reinterpretar y disputar los discursos y las emociones que conllevan, ni el fin de la lucha por el sentido de lo político, sino una

transformación en la manera en que las emociones se producen, circulan y utilizan para influir en la opinión pública.

En el plano metodológico, se propuso que el ACD, entendido como el estudio del discurso en tanto práctica social inscrita en relaciones de poder y en contextos concretos, ofrece un marco sólido para observar cómo las palabras, las imágenes y las formas de enunciación pueden reproducir o transformar jerarquías políticas y sociales. De manera complementaria, el enfoque del *framing* permite identificar qué aspectos de la realidad se resaltan, cómo se explican sus causas y responsabilidades y con qué carga emocional se proponen soluciones. Esta doble mirada se articula, además, con la tríada retórica *ethos-pathos-logos*, que ayuda a desentrañar la lógica persuasiva de los discursos en un escenario donde, el *ethos* (la credibilidad, la imagen del político) y el *pathos* (la emoción) han desplazado en buena medida al *logos* (la razón).

El aporte central de la propuesta es ofrecer una herramienta rigurosa y operativa para leer la política desde la dimensión emocional desde una perspectiva situada y crítica. Esto implica rastrear qué emociones se invocan, qué marcos de legitimidad las sostienen, qué identidades colectivas instituyen (nosotros/ellos) y qué futuros proyectan. Con ello, se supera la falsa dicotomía entre razón y emoción y se sitúa el foco en las mediaciones que hacen posible la adhesión, la identificación o el rechazo en democracias atravesadas por la lógica de la atención y la inmediatez. En suma, la propuesta permite analizar no sólo qué dicen los actores políticos, sino también cómo lo dicen, por qué lo dicen y qué emociones movilizan. El uso conjunto de estas herramientas muestra cómo las palabras y los signos no solo transmiten información, sino que también reconfiguran relaciones de poder, orientan la interpretación de la realidad y modelan disposiciones emocionales.

Referencias

- Ahmed, S. (2015). *La política cultural de las emociones*. Universidad Nacional Autónoma de México (UNAM).
- Aira, T. (2020). *La política de las emociones: Cómo los sentimientos gobiernan el mundo*. Arpa & Alfíl Editores.
- Arfuch, L. (2016). El "giro afectivo". Emociones, subjetividad y política. *deSignis*, 24, 245–254.
- Arias Maldonado, M. (2016). *La democracia sentimental: Política y emociones en el siglo XXI*. Página Indómita.
- Aristóteles. (2002). *Retórica* (A. Tovar, Trad.). Alianza Editorial. (Obra original publicada ca. 335 a.C.)
- Cabanas, E., & Illouz, E. (2019). *Happycracia: Cómo la ciencia y la industria de la felicidad controlan nuestras vidas*. Ediciones Paidós.
- Cantero, C. (2018). *Sociedad digital: Razón y emoción*. Editorial UOC.
- Charaudeau, P. (2021). *El discurso político: Las máscaras del poder*. Prometeo Libros.
- Chihu, A. (2012). La teoría del framing: Un paradigma interdisciplinario. *Acta Sociológica*, 59, 77–101.
- Chihu, A. (2024). *El «framing» de la comunicación política: Las maneras 2018-2022*. Gedisa.
- Colomer, J. M. (2023). El impacto de la globalización sobre la democracia. En *La humanidad amenazada. ¿Quién se hace cargo del futuro?* (pp. 45–62).
- Crouch, C. (2004). *Posdemocracia*. Taurus.
- Crouch, C. (2020). *Post-democracy after the crises*. Polity Press.
- Cvetkovich, A. (2018). *Un archivo de sentimientos: Trauma, sexualidad y culturas públicas lesbianas* (M. D. M. i Torrent, Trad.). Ediciones Bellaterra.
- Dagatti, M. (2017). Las emociones políticas. Un modelo discursivo de estudio. *Rétor*, 7(1), 40–72.
- Depetris-Chauvin, I. (2019). *Geografías afectivas: Desplazamientos, prácticas espaciales y formas de estar juntos en el cine de Argentina, Chile y Brasil (2002-2007)*. Latin American Research Commons.
- Fairclough, N., & Wodak, R. (2000). Análisis crítico del discurso. En T. A. van Dijk (Comp.), *El discurso como interacción social* (pp. 367–404). Gedisa.
- Flatley, J. (2008). *Affective mapping: Melancholia and the politics of modernism*. Harvard University Press.
- Fraser, N. (2023). *Capitalismo canibal: Qué hacer con este sistema que devora la democracia y el planeta, y hasta pone en peligro su propia existencia* (A. I. Bárcena, Trad.). Siglo XXI Editores.
- Gallardo, B. (2017). Pseudopolítica en la red: Indicadores discursivos de desideologización en Twitter. *Pragmalingüística*, 25, 189–210.

- Gantús, F., Rodríguez, G., & Salmerón, A. (Coords.). (2021). *El miedo: El más político de las emociones. Argentina y México, siglos XVIII-XX*. Instituto Mora/Universidad Autónoma de Zacatecas.
- Goffman, E. (2006). *Frame analysis: Los marcos de la experiencia* (C. P. Llovet, Trad.). Centro de Investigaciones Sociológicas (CIS).
- Gómez, A., & Velasco, G. (2024). *Atlas político de emociones*. Editorial Trotta.
- Guerrero, M. A. (2019). ¿El fin de la razón? La destrucción emocional de la democracia moderna: Propuesta desde la comunicación. *Signa: Revista de la Asociación Española de Semiótica*, 28, 1043–1065.
- Gutiérrez, S., & Plantin, C. (2010). Argumentar por medio de las emociones: La campaña del miedo del 2006. *Comunicación y Política*, 24, 41–69.
- Hall, S. (2004). Codificación y descodificación en el discurso televisivo. *CIC. Cuadernos de Información y Comunicación*, 9, 207–226.
- Illouz, E. (2007). *Intimidades congeladas: Las emociones en el capitalismo* (M. V. Rodríguez, Trad.). Katz Editores.
- Illouz, E. (2012). *Por qué duele el amor: Una explicación sociológica* (M. V. Rodríguez, Trad.). Katz Editores.
- Illouz, E. (2019). *Capitalismo, consumo y autenticidad: Las emociones como mercancía* (M. V. Rodríguez, Trad.). Katz Editores.
- Instituto Nacional Electoral (INE). (s. f.). *Voto y elecciones. Elecciones 2024*. <https://portal.ine.mx/voto-y-elecciones/elecciones-2024/>
- López, O. (2024). Los giros del giro afectivo: La centralidad de la vida sensible para teorizar lo social. Una lectura en clave latinoamericana. *Historia y Grafía*, 31(62), 263–301.
- Macón, C. (2013). *Sentimus ergo sumus*. El surgimiento del “giro afectivo” y su impacto sobre la filosofía política. *Revista Latinoamericana de Filosofía Política*, 2(6), 1–32.
- Macón, C., & Solana, M. (Coords.). (2015). *Pretérito indefinido: Afectos y emociones en las aproximaciones al pasado*. Título Editor.
- Maíz, R. (2010). La hazaña de la razón: La exclusión fundacional de las emociones en la teoría política moderna. *Revista de Estudios Políticos*, 149, 11–45.
- Moreno, A. (2024). Facetas de la polarización política en México. *Foro Internacional*, 64(2), 189–238.
- Mouffe, C. (1999). *El retorno de lo político: Comunidad, ciudadanía, pluralismo, democracia radical* (M. Lezcano, Trad.). Paidós.
- Mouffe, C. (2023). *El poder de los afectos en la política: Hacia una revolución democrática y verde* (J. I. S. y Gómez, Trad.). Siglo XXI Editores.
- Nussbaum, M. C. (2014). *Las emociones políticas: ¿Por qué el amor es importante para la justicia?* (M. V. Rodil, Trad.). Paidós.
- Pasquino, G. (2016). La personalización de la política: Más peligros que ventajas. *Cuadernos de Pensamiento Político*, 50, 19–24.
- Plantin, C. (2012). *La argumentación: Historia, teoría, perspectivas*. Biblos.

- Plantin, C. (2014). *Las buenas razones de las emociones*. Universidad Nacional de Moreno.
- Ramírez, I., & Falak, A. (2023). “Te amo, te odio: Dame más”. Polarización afectiva en la opinión pública argentina. *Revista SAAP. Sociedad Argentina de Análisis Político*, 17(2), 361–397.
- Scherer, K. R. (2024). Emotion processes in social and political contexts: The case of affective polarisation. *Cognition and Emotion*, 38(4), 411–417. <https://doi.org/10.1080/02699931.2024.2339570>
- Schröder, K. C. (2014). Realidades discursivas. En K. B. Jensen (Ed.), *La comunicación y los medios: Metodologías de investigación cualitativa y cuantitativa* (pp. 173–208). Fondo de Cultura Económica.
- Sibila, P. (2024). *Yo me lo merezco: De la vieja hipocresía a los nuevos cinismos*. Taurus.
- Sierra, M. (2015). Entre emociones y política: La historia cruzada de la virilidad romántica. *Rúbrica Contemporánea*, 4(7), 11–25.
- Siria, C. (2019). Emoción y razón en la construcción del discurso político: El caso de la proclamación de la no reconocida República Catalana en octubre de 2017. *Estudios de Lingüística del Español*, 40, 95–115.
- Van Aelst, P., Sheafer, T., & Stanyer, J. (2012). The personalization of mediated political communication: A review of concepts, operationalizations and key findings. *Journalism*, 13(2), 203–220. <https://doi.org/10.1177/1464884911427802>
- van Dijk, T. A. (2009). *Discurso y poder* (L. B. de F. de Aguilar, Trad.). Gedisa.
- van Dijk, T. A., & Rodrigo, I. (1999). *Análisis del discurso social y político*. Abya-Yala.
- Žižek, S. (2010). *En defensa de la intolerancia*. Sequitur.
- Žižek, S. (2024). *Argumentos: ¿Qué nos espera cuando no hay futuro?* (R. R. C. de Castro, Trad.). Anagrama.

Diversidad religiosa en América: Un breve recorrido por la hibridación, la resistencia y la revitalización

Religious diversity in the Americas: A brief overview of hybridization, resistance, and revitalization

Tatiana Cabrera Mogrovejo

tatyrytyk@gmail.com

Investigadora independiente

ORCID ID 0000-0003-2798-2457

Oswaldo Suin Guaraca

oswaldosuin@gmail.com

Investigador independiente

ORCID ID 0009-0008-7736-268X

Resumen: El manuscrito se centra en la complejidad y diversidad de las prácticas religiosas en América; con un especial énfasis en América Latina examina la religiosidad popular como resultado de la hibridación cultural entre influencias europeas, indígenas y africanas. Una sección significativa aborda la historia y la doctrina del catolicismo y su profunda influencia hegemónica en la región, y lo contrasta con el resurgimiento de las cosmovisiones indígenas y sus celebraciones. Asimismo, el estudio presenta un breve análisis de las religiones africanas, en el que destaca su sincretismo y la lucha contra la estigmatización racial, y concluye con una discusión sobre la creciente expansión del cristianismo protestante en el continente. El estudio aplica una metodología analítico-interpretativa con enfoque histórico, que examina el objeto de estudio a partir del análisis crítico de fuentes y su contextualización en los procesos históricos propios de la región.

Palabras clave: Diversidad religiosa; religiosidad popular; hibridación; sincretismo religioso; cosmovisiones.

Abstract: The manuscript focuses on the complexity and diversity of religious practices in the Americas; with a special emphasis on Latin America, it examines popular religiosity as a result of cultural hybridization between European, indigenous, and African influences. A significant section addresses the history and doctrine of Catholicism and its profound hegemonic influence in the region, contrasting it with the resurgence of indigenous worldviews and celebrations. The study also presents a brief analysis of African religions, highlighting their syncretism and the struggle against racial stigmatization, and concludes with a discussion of the growing expansion of Protestant Christianity on the continent. The study applies an analytical-interpretative methodology with a historical approach, examining the subject matter through critical analysis of sources and their contextualization within the historical processes specific to the region.

Keywords: Religious diversity; popular religiosity; hybridization; religious syncretism; worldviews.

Introducción

Hablar de la diversidad religiosa de América no es tarea fácil, si bien gran parte del continente americano; específicamente Latinoamérica, comparte un pasado histórico, éste tiene diferentes matices, por ende, hacia sus adentros circula una gran diversidad de prácticas, creencias espirituales y ritos, que sitúan a la religión como una esfera compleja y diversa. Esta heterogeneidad pone de manifiesto una serie de dinámicas en las que se entretajan diálogos interdisciplinarios e interreligiosos que se exteriorizan en una multiplicidad de creencias sobrenaturales que van desde las cosmovisiones indígenas, las religiones populares, las religiones afro, espiritualidades contemporáneas, entre otras. Las cuáles son impensables sin la presencia y la influencia católica,

pues la figura europea en América forzó el anclaje ideológico a la institución tradicional cristiana.

En este tenor, el presente trabajo tratará en una primera instancia realizar una aproximación a la religiosidad popular del contexto latinoamericano en sus rasgos generales, dado que por su gran riqueza y diversidad cultural es imposible alcanzar un estudio concreto y diferencial. Posteriormente se desarrollará la importancia de la presencia del catolicismo, como un elemento fundamental en la construcción histórica de Latinoamérica para luego regresar la mirada a las cosmovisiones indígenas y afros, y finalmente terminar con la presencia de la Iglesia cristiana en América.

Método

El presente estudio se fundamenta en una metodología de carácter cualitativo, basada en la revisión bibliográfica de fuentes académicas especializadas en ciencias sociales, historia, antropología y estudios religiosos. En este sentido, la naturaleza del estudio es analítica interpretativa con un enfoque histórico, pues busca la interpretación crítica de los fenómenos religiosos, tomado la dimensión histórica como un eje crucial para comprender la diversidad religiosa del continente americano.

Se privilegió la consulta de textos que abordan críticamente los procesos de hibridación cultural, las formas de sincretismo y las dinámicas de resistencia espiritual en América Latina, con especial atención a los estudios etnográficos que documentan las prácticas religiosas en contextos locales. Asimismo, se incorporaron investigaciones que analizan la expansión y hegemonía del catolicismo en el continente, así como trabajos que visibilizan la revitalización de las cosmovisiones indígenas y afrodescendientes. El corpus bibliográfico fue seleccionado bajo criterios de pertinencia temática, actualidad y rigurosidad académica, abarcando tanto autores clásicos como producciones contemporáneas.

La estrategia metodológica adaptada no pretende ofrecer un análisis exhaustivo de cada tradición religiosa o creencia, sino una aproximación

panorámica que permita comprender los principales vectores históricos y socioculturales que configuran la pluralidad religiosa en América. En este marco, el enfoque interpretativo empleado reconoce la religión como un fenómeno dinámico, profundamente entrelazado con las estructuras de poder, identidad y memoria colectiva. La revisión bibliográfica fue sistematizada en función de los ejes temáticos del manuscrito: religiosidad popular, catolicismo hegemónico, cosmovisiones indígenas, religiones afrodescendientes y expansión del protestantismo. Esta organización permitió establecer conexiones analíticas entre los distintos procesos religiosos, destacando tanto sus tensiones como sus formas de resignificación y continuidad.

La delimitación geográfica del estudio se centró principalmente en el continente de América, pero con un énfasis especial en América Latina. La metodología utilizada priorizó la revisión de textos que analizan los procesos de sincretismo y resistencia espiritual en el contexto latinoamericano. Específicamente, se incluyeron datos y análisis de países como Brasil, Ecuador, y otras naciones sudamericanas y centroamericanas donde las religiones afrolatinas y el protestantismo han cobrado fuerza. Asimismo, se incorporó el análisis de la presencia del cristianismo en Norteamérica, particularmente en Estados Unidos, abordando su fundamentalismo y la expansión de la Iglesia cristiana en el continente. En este tenor, el estudio ofrece una aproximación panorámica de la pluralidad religiosa en toda América.

El manuscrito abarca un extenso marco temporal, iniciando con la conquista y posterior colonización de América Latina, periodo en el que se impuso el catolicismo y se configuró la religiosidad popular a través de la hibridación. Continúa con la época de la hegemonía católica y se extiende hasta el presente, cubriendo el resurgimiento de las cosmovisiones indígenas a partir de los años 90, y la expansión del protestantismo pentecostal en las dos últimas décadas. Finalmente se incorporaron datos y fenómenos religiosos de años muy recientes.

La religiosidad popular latinoamericana

La religión se enmarca en un cerco conceptual de gran complejidad, de ahí que, existan serios problemas para definirla sin caer en un religiocentrismo. En este sentido, el término de “religión popular” no ha escapado de esta problemática conceptual, pues tiende a enredarse en varias definiciones y muchos usos que terminan siendo discordantes. Sumado a ello, la religión popular suele tener un carácter despectivo cuando se la mira como una religión degradada de la religión institucional, situándola como una religión inferior o inculta (De la Torre, 2021).

No obstante, otras miradas de la religión popular, se refieren a ella como un “campo religioso propio y con autonomía relativa, que tiene por sujeto al pueblo, aunque inciden sobre él sacerdotes, samanes y profetas” (Dussel, 1986, p. 105). Asimismo, la religiosidad popular ha sido vista como una religiosidad salvadora de las opresiones del pueblo (De la Torre & Martín, 2016), y es justamente en este marco que hacen su aparición la teología de la liberación y las “Comunidades Eclesiales de Base” como una reivindicación del pueblo a la consumación profética para alcanzar la soberanía de Dios en la tierra, pero ahora alineada a ideas propias de la modernidad como la igualdad, la justicia social y la libertad (Dussel, 1986).

Es bien sabido que con la conquista y posterior colonización de América Latina se impuso el catolicismo como una religión oficial, por lo que las religiosidades y sus prácticas se configuraron por una influencia del toma-y-da entre europeos, indígenas, y en parte, por los afroamericanos (Brosseder, 2018), es decir, existió un cambio en ambas direcciones. En palabras de Estermann (2003), no se produjo solo una asimilación de una parte, sino una relación inter – transcultural entre ambas religiones, por lo tanto, el autor no apunta a un “sincretismo andino” resultado del contacto entre culturas sino producto de la interacción y el diálogo, a pesar de la persecución y extirpación de que los europeos llamaron idolatrías. Por ende, la actual religiosidad popular en Latinoamérica es fruto de una hibridación de prácticas culturales occidentales, indígenas y afros, dicho por Renné de la Torre:

La religiosidad popular es un rasgo característico de América Latina no solo por su colorido y vivacidad, sino sobre todo por ser una encrucijada histórica que mantiene la tensión del mestizaje cultural que ha tenido lugar en distintas etapas de la vida de este subcontinente. Este mestizaje resultó de un violento choque cultural y surgió de la imposición del catolicismo europeo y la resistencia indígena; y que posteriormente cohabitó con las religiones africanas. (2021, p. 259)

El proceso intercultural e interreligioso entre el catolicismo y las diversas religiosidades de los contextos de Latinoamérica quedan claramente evidenciados en las múltiples manifestaciones religiosas. Un ejemplo de ello es el barroco religioso de los Andes, con todos los rituales, festividades y oralidades que giran en torno a éste (Ríos, 2008). Elementos culturales propios de los pueblos indígenas, y también de los afros, han encontrado una resistencia a través del tiempo al conservar conocimientos y saberes que han perdurado al proceso de colonización y que hoy se hacen evidentes dentro de sus ritos y festividades religiosas. Así podemos mencionar el Corpus Christi; una de las festividades de mayor transcendencia en el calendario litúrgico católico, donde el mestizaje religioso del que habla Estermann (1998), se hace evidente en los elementos andinos cristianizados y en los elementos cristianos andinizados, dado que “el Corpus Christi que se festeja actualmente, se fundamenta en la cosmovisión andina, en sus fiestas ancestrales, primero del maíz y luego del Inti Raymi. Es más, la festividad católica coincide con la época de maduración del maíz” (Morocho, 2008, p. 103).

Dentro de este contexto de hibridación, la Iglesia Católica ha mantenido la institucionalización y jerarquía de sus prácticas y creencias, sin embargo, la religiosidad popular sin la necesidad de ir en contra de la autoridad del catolicismo, ha abierto camino a expresiones propias que no pudieron ser suprimidas por el colonialismo a pesar de todos sus intentos por extirparlas. De ahí que, Marzal (2005) señale que el cristianismo ha encontrado el modo de cohabitar durante milenios con las cosmovisiones nativas americanas consentidas bajo el oficialismo de los cristos, vírgenes y santos, y sus mediaciones milagrosas. Incluso muchas de estas

manifestaciones de religiosidad popular hoy en día son resguardadas como patrimonio cultural material e inmaterial como un intento de resguardar la cohesión social por medio de la identidad y la memoria. Empero, bajo la premisa de que la cultura es dinámica, todas estas prácticas de religiosidad están en constante cambio y resignificación.

La presencia del catolicismo en América

La iglesia católica es la institución religiosa más antigua establecida en América desde la llegada de los europeos. Su influencia ha sido tan grande, que no se puede analizar la historia de nuestro continente sin hacer referencia a su hegemonía y control sobre las instituciones sociales, solo basta con recordar los inicios de los Estados – Naciones en Latinoamérica y su adscripción al catolicismo como religión oficial de las nacientes repúblicas (Ramírez, 2009).

Según los atributos propios del catolicismo, su iglesia se establece bajo cuatro principios: Una, Santa, Católica y Apostólica. Por lo que sus fieles profesan su fe mediante el Credo de los Apóstoles y el Credo Niceno-Constantinopolitano. La Iglesia Católica es “Una” ya que representa a Dios mismo y Dios es uno según su propia doctrina, así San Pablo en su Primera Carta a los Corintios se refiere a la Iglesia como “El Cuerpo de Cristo”. En cuanto al atributo de la Santidad, la Iglesia católica se dice Santa pues “Santo” es su fundador y “santos” son sus fines y objetivos, es decir, contiene la plenitud de los medios de santificación y salvación, por lo tanto, es Santa porque sus seguidores están llamados a ser santos.

Por otro lado, la Iglesia argumenta ser católica traducida como “universal” porque Cristo está presente en ella, en cuanto pretende anunciar la Buena Nueva y acoger en sus adentros a toda la humanidad, de cualquier lugar y en todo momento a través del bautismo como uno de los medios de salvación, siendo el sacerdote el guía espiritual de la Iglesia y el encargado de ejecutar todos los sacramentos indispensables para lograr la santidad (Trigilio & Brighenti, 2007). Es apostólica, por sus cimientos en Pedro y los otros apóstoles, y porque las enseñanzas de éstos forman parte de los principios del

catolicismo. Así, la misma autoridad de la que gozaron los apóstoles en los tiempos de Cristo, hoy está en manos del Papa, los Obispos y los sacerdotes, siendo el primero la autoridad absoluta de la Iglesia Católica como vicario de Cristo en la tierra. La Iglesia Católica se dice heredera de la tradición y la doctrina fundada por Jesucristo, hijo de Dios hecho carne, y se considera a sí misma como la única y legítima representante de Cristo (Alves, 2011).

La doctrina fundamental del catolicismo se encuentra en el Credo, donde se ubican las explicaciones sistemáticas del catecismo. Del mismo modo, el Credo contiene todas las cláusulas de fe erigidas en los primeros concilios de la historia (Alves, 2011). Por su parte la Revelación también constituye parte trascendental de la doctrina católica, pues en ella confluyen la Sagrada Escritura y la Tradición. Para los creyentes católicos el culmen de la Revelación es Jesucristo (Vatican, 2012). El Papa, considerado como sucesor de San Pedro y máxima autoridad de la Iglesia Católica, es el mayor referente de la doctrina católica, su figura ha variado a lo largo de los siglos y es elegido por los cardenales en lo que se denomina conclave (Trigilio & Brighenti, 2007).

Dentro del catolicismo es importante la creencia en los santos y en las vírgenes, siendo la veneración más significativa hacia la Inmaculada Concepción, es decir, La Virgen María, madre de Jesús. En sí, la Iglesia Católica dota a estos personajes de un poder espiritual capaz de absolver los pecados mediante los rezos y las penitencias. Otra de las características que guarda esta religión es la creencia de la presencia real de Jesucristo en el ritual de la Eucaristía, cristalizado por medio de La misa, donde el pan y el vino se transfiguran en el cuerpo y sangre de Cristo (Ratzinger, 1995), diálogo directo entre el ser humano y el ser divino. La misa, como ritual religioso, se estructura en diferentes momentos, destacando entre ellos la liturgia y la comunión con Dios. Ésta es dirigida por el sacerdote y se celebra todos los domingos, día considerado de descanso para los católicos, ya que representa la resurrección de Jesucristo. Para alcanzar la santidad, la vida después de la muerte, se requiere de 7 sacramentos: sacramentos de iniciación cristiana (Bautismo, Eucaristía y Confirmación), sacramentos de curación (penitencia

y unción de los enfermos) y los sacramentos de servicio (matrimonio y orden sagrado) (Vatican, 2012).

Las festividades ligadas al catolicismo se enmarcan dentro del calendario litúrgico de su iglesia, en el que destacan el Adviento, la Navidad, la Cuaresma y la Pascua, considerados como tiempos “fuertes” del año litúrgico, en los que los fieles participan con mayor presencia y devoción. Sin embargo, como se evidenció en líneas anteriores al hacer referencia a la religiosidad popular en América Latina, cada una de estas festividades guarda una hibridación cultural que en la praxis se aleja de la ortodoxia católica por medio de manifestaciones autóctonas de su contexto. Así, la fiesta religiosa, no solo cumple con el ritual católico, sino que actúa como un cohesionador de la vida social en nuestro continente (Aguirre, 2020) por lo que el rito no termina en la Iglesia, sino que se traslada a la plaza o a las calles, donde cobran vida las procesiones, las romerías, las danzas, los juegos populares, entre otros. De ahí que, no sea extraño que cada una de las ciudades, departamentos, cantones y parroquias de Latinoamérica tenga su propio santo patrono o virgen, a quienes les rinden culto cada año por medio de grandes fiestas llenas de parafernalias.

Varias de estas devociones fueron introducidas en América Latina por las congregaciones religiosas durante la época colonial, quienes jugaron un papel central no solo en la evangelización, sino en la educación, pues las primeras escuelas, colegios y universidades fundadas en los territorios conquistados estuvieron en manos de las congregaciones como los jesuitas, los dominicos, los franciscanos, los salesianos, las clarisas pobres, entre otras. El objetivo de estas congregaciones se orienta a anunciar la Buena Nueva de Dios en el mundo y cada una de ellas se desarrolla bajo los criterios establecidos por su fundador. Los miembros de las órdenes religiosas y las congregaciones viven en comunidad y profesan votos que dependen de la orden a la que pertenezcan (Trigilio & Brighenti, 2007).

Bajo la estructura del catolicismo descrita, miles de fieles se adscriben a sus fundamentos como un modo de vida, así actualmente encontramos en América el 48% del total de católicos en el presente mundo, de los cuales el

57% reside en América del Sur y el 27% solo en Brasil, lo que lo convierte en el país con mayor número de bautizados católicos en el mundo, quienes representan cerca de 180 millones de personas (Vatican News, 2023). En Ecuador el 74,8% de su población profesa el catolicismo (INEC, 2022). Estos datos revelan que, pese la enorme diversidad cultural en América, la religión cristiana católica se consolidó como el eje articulador de la vida social, especialmente en América Latina, lo que configura un marco religioso dominante frente al cual dialogan prácticas y cosmovisiones indígenas.

De regreso a las cosmovisiones indígenas

Los pueblos de América Latina comparten una historia similar en cuanto al colonialismo y a la conformación de repúblicas. Los Estados – nación se edificaron desde una lógica occidental y monocultural, que excluía la heterogeneidad de su población, sus prácticas culturales y su cosmovisión. Lo que desembocó en la negación, discriminación y exclusión de los pueblos indígenas y afros (Flores, 2022).

A partir de esta realidad, los indígenas y los afrodescendientes, han tratado de reivindicar sus derechos como ciudadanos mediante varias sublevaciones y levantamientos, cuyas demandas se pueden sintetizar en: la plurinacionalidad, la interculturalidad y la autodefinición como pueblos y nacionalidades. Demandas que, en países como Ecuador y Bolivia, terminaron inscritas dentro de sus respectivas Constituciones, pero que al momento de salir de la abstracción a la práctica se han encontrado con serias problemáticas, como el fuerte racismo estructural que se mantiene hasta la actualidad. Sea como fuere, a partir de los años 90, cuando se hacen más fuertes y notorias las sublevaciones indígenas, germinan con ellas el resurgimiento del capital simbólico cultural de las nacionalidades y pueblos indígenas como una forma de resistencia, incorporada en las celebraciones que se desprenden de su modo de ver el mundo.

Surge de este modo, el concepto de cosmovisión para referirse al pensamiento indígena y sus concepciones del entendimiento de la realidad y su interacción con ella (Simbaña & Caral, 2022). Concepto que no ha estado

libre de críticas eurocentristas, al ubicarlo como un término que se usa en oposición al conocimiento científico y lo reduce a meras creencias y supersticiones. No obstante, como lo señalan Simbaña y Caral, “la producción de conocimientos en cuanto a la práctica especializada y desarrollada por especialistas hombres y mujeres indígenas está impregnada de significados en los que subyacen concepciones cosmológicas y cosmogónicas de los pueblos indígenas” (2022, p. 85).

Siguiendo con la sabiduría que han resguardado los pueblos indígenas, emerge como pieza clave en la cosmovisión andina el concepto de Pacha, que puede entenderse como “espacio – tiempo” o viceversa como “tiempo – espacio” (Arriagada, 2019), así el espacio es referido a la materia y el tiempo al espíritu:

En el sentido estrecho, Pacha significa todo lo que es aprehensible. Lo que es visible, distinguible y adjetivable. En el sentido amplio, Pacha, significa alteridad, todo lo aleatorio, desconocido, indeterminado, como cuando las cosas vienen al mundo, los hechos, los eventos [...]. (Untoja Choque & Mamani Espejo, 2000, p. 9)

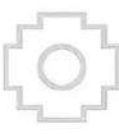
Así, Pacha entendida como la incorporación del espacio-(materia) y tiempo-(espíritu) es el origen de todas las cosas; generador de vida que está en constante cambio; la justa vinculación entre lo material y lo inmaterial del espacio – tiempo. La existencia del pensamiento andino se materializa a los adentros del Pacha donde las ideas toman forma ante una unidad de forma – idea o idea – forma, así ante la necesaria vinculación entre el concepto y la forma, la idea de espacio se relaciona y se expresa formalmente en la vida práctica por medio de un cuadrado o cubo, mientras que la idea del tiempo está representada por un círculo, un tiempo que retorna cíclicamente (Arriagada, 2019). La máxima del concepto de Pacha se origina en la constelación de la Chakana o “Cruz del Sur”, constelación que sirvió como eje orientador de la cosmovisión de las culturas andinas, y que sentó sus bases en dualidades complementarias denominadas Hanan (masculino-cuadrado) y Hurin (femenino-círculo), opuestos complementarios que se manifiestan en todos los aspectos de la vida misma (Chambi, 2009).

Estos opuestos complementarios también son aplicables a las divinidades andinas, así Wiracocha Pachayachachic, el gran ordenador del universo se expresa como una divinidad dual compuesta por Wiracocha y Pachacama como manifestaciones del orden y energías universales (Milla Euribe, 1990). Así, el espacio andino denominado *Tawa Runa Pacha* está dividido en cuatro mundos: *Hanan Pacha*, el mundo superior; *Kay Pacha*, el mundo de aquí; *Uku Pacha*, el mundo inferior; y, *Hawa Pacha*, el mundo de afuera (Simbaña & Caral, 2022). Y regido por los principios de la cosmovisión andina: complementariedad, reciprocidad, correspondencia y relacionalidad.

Bajo este contexto, los pueblos indígenas retoman la celebración de los *raymicuna* como una parte fundamental de la espiritualidad andina y como una posición epistémica desde el Sur alejada de posturas eurocéntricas y occidentales (Japón et al., 2022). Los *raymicuna* se desprenden del conocimiento de la *Chakana*, la que marca los solsticios y los equinoccios, indispensables para la siembra y la cosecha, por lo que toda festividad es acomodada al calendario agrícola. En este contexto, los pueblos andinos poseen cuatro celebraciones importantes: *Inti Raymi*, *Kulla Raymi*, *Kapak Raymi* y *Pawkar Raymi* (Flores, 2022); cada uno de ellos con sus características propias como lo muestra la figura 1.

Figura 1

Calendario ritual andino de los Raymi

					
<i>Fiesta a la Fecundidad de la madre tierra. La producción</i>	<i>Culto a la fertilidad de la madre tierra.</i>	<i>Fiesta Real, dedicado a la germinación</i>	<i>Fiesta a la Pacha Mama, época del florecimiento.</i>		
Inti Raymi	Kulla Raymi	Kapak Raymi	Pawkar Raymi		
Solsticio de invierno	Equinoccio de primavera	Solsticio de verano	Equinoccio de otoño		
21 de junio	21 de septiembre	21 de diciembre	21 de marzo		

Nota. CIOFF Ecuador, 2019.

Todas estas festividades están dirigidas por un tatita o una mama que desarrolla los rituales según el Raymi que corresponda. A lo que es necesario recalcar que estas festividades también han sufrido cambios a lo largo del tiempo y de ninguna manera son manifestaciones que han guardado su originalidad, pues si solo tomamos como ejemplo los *raymicuna*, se sabe que éstos eran celebrados por los incas más no por todos los pueblos existentes antes de la llegada de los europeos, los incas como un pueblo en expansión incorporaron estas festividades en las localidades conquistadas. Por lo que es importante dejar por sentado que no todas las comunidades indígenas celebran de manera homogénea las festividades ligadas a su cosmovisión. No obstante, si la lógica andina se ha mantenido, ha sido en parte por la capacidad de los pueblos indígenas de adaptar sus saberes y rituales a constantes transformaciones en la búsqueda de la reivindicación de sus derechos.

Religiones afrolatinoamericanas

En relación a la cosmovisión de los afrodescendientes en América, el análisis de las estructuras elementales de su actual conocimiento, nos lleva a trasladarnos a su adaptación histórica acontecida en los hechos de su

secuestro y conducción como esclavos a tierras americanas, dicho por Zapata Olivella:

En el proceso de adaptación a la sociedad esclavista, el africano, según encontrará valores o no identificables a sus cultura, sufrió una reafricanización o desafricanización de sus propios valores hasta poder desarrollar un nuevo tipo de sociedad negra que respondiera a las nuevas exigencias del medio. (1989, p. 102)

Hoy en día las religiones afrolatinoamericanas se ven envueltas en dos direcciones, por un lado, al igual que sucedió con los pueblos indígenas, existe una revalorización de sus saberes y conocimientos como respuesta a la exigencia del goce pleno de sus derechos como ciudadanos y ciudadanas; y, por otra parte, producto del colonialismo, el racismo ha llevado a que las religiones afrolatinas sean criminalizadas e incluso sufran de ataques de odio. Sin embargo, en el caso de las religiones afrocubanas no ha existido tal proceso de reivindicación, dado que, desde muy temprano su práctica trascendió fronteras y gozaron de su respectiva valorización, sobre todo en el caso de la Santería e *Ifá* o *Regla de Ocha-Ifá* de raíces yorubas, que incluso han sido reconocidas como patrimonio cultural del Cuba (Matory, 1998).

Por su historia, Brasil es el país de América, donde existe una mayor presencia de variantes de religiosidades de origen africano, no obstante, en las últimas décadas el número de seguidores se mantiene o apenas se eleva frente a la creciente oposición de los pentecostales y neopentecostales, quienes han adquirido un gran poder en todas las esferas, sobre todo en la política. Lo que ha dado origen; después de la persecución por parte de la Iglesia Católica, a una nueva ola de persecución de las religiones afros (Frigerio et al., 2022). En tanto que en Brasil estas religiones son atacadas, su presencia es cada vez más importante en países como Uruguay, Argentina, México, Venezuela, Colombia, Perú y Bolivia, aunque no están exentas de polémicas y estigmatizaciones:

Las fiestas a *Iemanjá*, una de las dedidades (orisha) más populares en América, se celebran por todo el territorio nacional y constituyen su

faceta socialmente mejor recibida (Frigerio). Los sacrificios de animales y ofrendas en lugares públicos, por otro lado, son los aspectos mediáticamente más controvertidos. Especialmente en Argentina, donde su aparición en la sección de notas policiales de los diarios es frecuente, debido a la asociación entre sacrificios de animales y posibles crímenes rituales. Aunque esto no es excepcional del Cono Sur, pues las religiones afrocubanas como la santería y el palo monte, en países como México y Colombia, también son referencia común en las secciones de justicia y delincuencia en diversos medios de comunicación. Sus representaciones históricas, asociadas a lo demoníaco, salvaje y primitivo, ahora se potencian y retroalimentan con la narrativa mediática asociada al crimen, violencia y narcotráfico. (Frigerio et al., 2022, p. 13)

Umbanda, candomblé y batuque (variantes de religiones afrobrasileñas), y la Santería e *Ifá* o *Regla de Ocha* (religiones afrocubanas), han encontrado cabida en países europeos como España, Alemania y Francia gracias a procesos de migración (Rossbach de Olmos, 2014). En el norte de América, específicamente en Estados Unidos, las religiones afrocubanas se han extendido mucho más que las afrobrasileñas. Del mismo modo, el vudú haitiano dejó de ser una religión practicada únicamente por los inmigrantes para establecerse en ciudades como Nueva Orleans (Dorsman, 2019).

En el caso de Ecuador, la religión de los afroecuatorianos; quienes representan un 7,8% de la población total (INEC, 2022), se compone de una mezcla de varios elementos provenientes de las religiones africanas, el catolicismo y la cosmovisión indígena, donde destacan ritos como los arrullos, los chigualos, los alumbramientos, los velorios, los gualíes, entre otros. Como lo menciona Antón “en esta interacción de tradiciones, el mundo africano se acomoda al mundo católico [...] De este modo, la celebración de un santo, de una virgen se convierten en una síntesis de la vida religiosa popular de los afrodescendientes” (2014, p. 78). En este sentido, es importante hacer notar, que gran parte de la población afro que llegó a Ecuador durante la época colonial, no fue esclavizada, fueron negros libres

que se asentaron en lo que hoy es Esmeraldas, donde además de formar palenques, establecieron alianzas con las poblaciones indígenas con las cuales compartían territorio. En suma, la religiosidad afroecuatoriana surge de un entramado histórico y cultural híbrido cuya complejidad interna fundamenta la riqueza de su universo cosmológico.

Norman Whitten (1992), en Antón (2014), identifica tres mundos en el universo cosmogónico de la religiosidad afroecuatoriana:

El mundo divino o de arriba (cielo), donde viven las deidades o fuerzas espirituales como los orishas, santos, vírgenes, espíritus, ángeles, arcángeles, San Pedro y Dios. El mundo humano o de abajo (limbo, purgatorio e infierno), en el que se encuentran las almas de los muertos, los diablos y Lucifer. El mundo terrenal (mar, ríos y montes), habitado por hombres y mujeres que interactúan con las fuerzas de la naturaleza, y donde habitan los seres mitológicos de la cultura afro, como la Tundra, el Riviel, la Gualgura, entre otras entidades. Las almas pueden viajar de un mundo a otro por medio de puertas de entrada, como son los cementerios. (p. 68)

De regreso a las religiones afrobrasileñas, nos aproximaremos a una de sus vertientes más comunes que son los *orixás* del complejo religioso del candomblé; grupo que se caracteriza por rendir culto a entidades espirituales que son divinidades de origen africano. Bajo sus principios toda persona es concebida como la instancia humana de una divinidad o un santo; por lo que cada visitante que ingresa a una casa de culto es analizado y los miembros de la comunidad intercambian una serie de comentarios respecto a quién de los *orixás* parece ser el dueño de su cabeza, en otras palabras, su santo protector, aquel que va con su personalidad (Dantas, 2011). Entonces la persona pasa a estar integrada por dos polos ontológicos, el yo inmediato y el yo metafísico (*orixá*), el primero es el responsable del aspecto moral, el comportamiento y la conciencia ética; y el segundo es el responsable de las opciones más decisivas de la vida que le dan su orientación reconocible y su estilo consistente (Segato, 1993). Una vez que por medio del rito de iniciación se le

asigna a la persona su *orixá*, ésta entra en la comunidad de culto conocida como familia de santo liderada por una madre o padre santo.

Dentro de esta ritualidad surgen varios componentes, como el lugar de culto denominado *terreiro*, espacio donde los conocimientos son afirmados por las prácticas del día a día, tal como lo indica María Ignacia Dávila (2009) cuando menciona que, las *madres de los santos* en los *terreiros* ocupan la posición de *Ialorixá*, pues al ser elegidas por sus divinidades, valoran en gran medida su ascendencia y linaje, y continúan así su herencia africana.

Si bien aún quedan muchas variantes de las religiones afroamericanas por explorar, su análisis va mucho más allá de los límites del presente estudio, debido a la complejidad que abarca el tema de la religiosidad. Sin embargo, se ha intentado realizar un acercamiento a las religiones que comprenden una mayor población, y a las que los autores de este manuscrito tienen una mayor cercanía. No obstante, hemos tratado de no caer en generalizaciones, pues son pocos los aspectos de culto que se pueden englobar. Finalmente, para culminar este apartado es preciso insistir que las religiones que tiene su origen en el continente africano aún conservan una fuerte estigmatización debido al racismo estructural heredado del colonialismo, lo que las ha llevado a vincularlas negativamente con la hechicería, la magia y las supersticiones. A pesar de ello la reivindicación tanto cultural como racial que destaca las raíces negras de estas religiones se ha convertido en un estandarte poderoso de legitimación.

El cristianismo en EE.UU. y Latinoamérica

La época colonial en Norteamérica estuvo marcada por la religión cristiana y La Biblia, convertida en la única fuente de revelación, desde la cual surge el mito inglés que concebía a Inglaterra como el Nuevo Israel (la nación elegida) y al peregrinaje puritano del Nuevo Mundo como una proeza del pueblo elegido por Dios para fundar en Norteamérica la tierra prometida, el nuevo Canaán (Stark, 2012). De ahí que, Norteamérica proclame su libertad de la mano de la religión como la salvaguardia de sus costumbres y a las costumbres como garantía de las leyes.

Esta visión histórica del fundamentalismo marcó a EE.UU. como una nación protegida y bendecida por Dios. Sin embargo, el modernismo vino a confrontar este fundamentalismo, y aparentemente salió victorioso hasta los años 80 del siglo XX donde una nueva generación compuesta por varias organizaciones formó una nueva derecha cristiana con el objetivo de lograr una recristianización (Gullo, 1997), vista como una reconquista de América. Tales organizaciones cristianas si bien guardan las mismas creencias son independientes entre sí, empero, cuando se trata de unir sus esfuerzos en situaciones como la prohibición del aborto, el restablecimiento de la oración en las escuelas, la promoción de leyes en las escuelas para enseñar el creacionismo en igualdad con el evolucionismo; la vinculación entre estas entidades es inmediata. Todo ello, explica el actual “nacionalismo cristiano” que se observa en EE.UU., donde la cultura evangélica tiene una doble mirada, tanto religiosa como política.

En el caso de América Latina, en las últimas décadas el protestantismo se ha expandido en este continente, sobre todo los cristianos pentecostales, cuya religión plantea una experimentación de los dones del espíritu santo mediante “la sanación divina, hablar en lenguas y recibir revelaciones directas de Dios, sin la intervención de un tercero. Y, a diferencia de la Iglesia católica, no hay una autoridad central de poder” (Ceriani, 2013, p. 18). El pentecostalismo, inició su ascenso en Latinoamérica a partir de los años 60, pero ha sido en estas dos últimas décadas que ha ganado un mayor protagonismo; así cientos de católicos que se han convertido a esta rama del cristianismo, han encontrado una iglesia que le da mayor importancia a la moral en la vida cotidiana, y que abraza el individualismo y el progreso. Esto se refleja incluso en la política, pues varios líderes políticos, como Bolsonaro, han gobernado con los principios cristianos de la mano.

El pentecostalismo y el evangelismo se ha consolidado también en las poblaciones indígenas, por ejemplo, en Ecuador existe una importante influencia de la religión cristiana (evangélica) en las comunidades indígenas de los Andes, sobre todo en la Sierra centro sur del país, donde por ubicarse la mayor población evangélica, las organizaciones cristianas extranjeras

promueven la ayuda internacional a través de proyectos que benefician a la comunidad con el fin único de adherir más creyentes a su iglesia (Viteri, 2012).

Según la Encuesta de Opinión Pública de CID Gallup (2022), a pesar que Latinoamérica continúa siendo eminentemente católica, existen un significativo número de evangélicos que se distribuyen de la siguiente manera: Honduras 44%, Guatemala 43%, El Salvador 40%, Nicaragua 38%, Panamá 31%, República Dominicana 26%, Costa Rica 25%, Venezuela 23%, Perú 20%, Ecuador 20%, Colombia 20%, México 20% y Argentina 10%. En Estados Unidos la religión más extendida es el cristianismo con un 74, 54%.

Los principios del cristianismo están basados en la “Declaración de verdades fundamentales”, siendo el bautismo en el Espíritu Santo, la sanidad divina, y la segunda venida de Cristo los fundamentos más importantes de la iglesia. A diferencia del catolicismo, el cristianismo es mucho más simple en el sentido de que no adora santos y vírgenes. Además. profesa el sacerdocio universal, que implica la libre interpretación de las escrituras. Su estructura es descentralizada y no reconocen al Papa como autoridad. Del cristianismo han resultado varias iglesias como la católica, la ortodoxa, la anglicana, la calvinista, la copta, la iglesia luterana y los protestantes en general, cada una de ellas con sus particularidades.

Conclusiones

Ahondar en temas religiosos es una tarea bastante complicada, pues involucra un análisis muy profundo que suele ser muy difuso si hacemos referencia a las diversidades de creencias, saberes, filosofías, cosmovisiones, etc., existentes en el mundo. Es decir, encontrar un punto de partida para la explorar el sentido religioso sin caer en religiocentrismos es una labor muy ardua. Sin embargo, el manuscrito presentado trató de realizar una aproximación a la diversidad de religiones de América, quizá no de una manera minuciosa, pero si enmarcado en las prácticas de la mayoría de su población, lo cual no significa que las religiones que no se desarrollaron carezcan de importancia, más bien obedeció a delimitar la extensión de este trabajo.

Las religiones tradicionales en América, entendidas como las que se han transmitido de generación en generación en este territorio, se practican en función al contexto histórico del continente, el que presenta una gran diversidad de pueblos y nacionalidades, y por ende de religiones. Destaca la religión popular, producto del encuentro cultural entre pueblos indígenas, afros y europeos, cuyas expresiones religiosas – como la festividad litúrgica del Corpus Christi – se cristalizan en los ritos y festividades mediante una hibridación cultural, dando como resultado un catolicismo que confluye con las lógicas andinas y africanas. Esta religión popular que resistió los intentos de extirpación del colonialismo, ha abierto camino a expresiones propias y está en constante cambio y resignificación.

Las cosmovisiones indígenas y las religiones de los afrodescendientes representan esferas de resistencia espiritual y han sido fundamentales para el regreso del sentido de dignidad a grupos históricamente excluidos y perseguidos. Así, los pueblos indígenas retomaron celebraciones como los *raymicuna* como una posición epistémica no eurocéntrica, ajustada al calendario agrícola. Por su parte las religiones afrolatinoamericanas, resultado de la diáspora africana, han encontrado en la reivindicación una alternativa para mostrar su lucha en contra de las prácticas coloniales que han llevado a estigmatizar y perseguir sus creencias, reduciéndolas a supersticiones y brujerías. Fenómeno que no ha impedido que sigan creciendo no solo a nivel de América Latina, sino en Europa producto de la migración.

Históricamente, el catolicismo ha mantenido su posición como la institución religiosa más fuerte y con mayor número de fieles en la región. Su hegemonía y control han sido tan profundas que la historia del continente americano es impensable sin su referencia. No obstante, en las últimas décadas ha crecido significativamente la expansión de las religiones protestantes, especialmente en las vertientes pentecostales que se originaron en Estados Unidos. El crecimiento protestante es notable por su vinculación con líderes políticos en América Latina, lo que ha llevado a la implantación de un tipo de “nacionalismo cristiano”, similar al observado en Estados Unidos.

Finalmente, es preciso mencionar que todas las religiones están en un constante cambio al igual que las culturas, muchas de ellas se han ido adaptando a una serie de transformaciones que son importantes para mantener la cohesión social. En otros casos, como las cosmovisiones indígenas y las religiones de los afrodescendientes, éstas han otorgado un sentido de regreso de la dignidad de los grupos que fueron históricamente excluidos y perseguidos. Lo que ha hilado muy fino, pues muchos de estos pueblos en la búsqueda de reivindicación de sus derechos han caído en esencialismos. A pesar de ello, toda la diversidad de religiones que se practican en América está en un continuo dinamismo y cambio, pero lejos de desaparecer, de ahí que, la cotidianidad sociocultural de nuestro continente no puede ser entendida fuera de la esfera religiosa.

Referencias

- Aguirre-Romero, F. (2020). Tradición y transmisión de la fe. El caso de la «religiosidad popular» en el Chile actual. *Scripta Theologica*, 52(1), 215-243.
- Alves, T. (2011). *El catolicismo* [Tesis de grado, Universidad Adventista de Bolivia].
- Antón, J. (2014). *Religiosidad afroecuatoriana*. Instituto Nacional de Patrimonio Cultural (INPC).
- Arriagada Peters, L. (2019). Avatares de la forma en el espacio-tiempo Pacha. *Tópicos del Seminario*, 42, 165-204.
- Brosseder, C. (2018). *El poder de las Huacas*. Ediciones El Lector.
- Ceriani, C. (2013). La religión como categoría social: encrucijadas semánticas y pragmáticas. *Antropología Social y Cultural*, 7(1), 10-29.
- Chambi Pumakahua, O. (2009). *Uyupanas y Kipus numéricos. Operaciones aritméticas en el Sistema Uyupana-Kipu*. Universidad de Tarapacá (UTA).
- De la Torre, R. (2021). La religiosidad popular de América Latina: una bisagra para colocar *lived religion* en proyectos de descolonización. *Revista Cultura & Religión*, 15(1), 259-298.
- De la Torre, R., & Martín, E. (2016). Religious studies in Latin America. *Annual Review of Sociology*, 42, 473-492. <https://doi.org/10.1146/annurev-soc-081715-074427>
- Dorsman, R. (2019). The Spirit(s) of New Orleans: Community healing through commemoration. En M. Balkenhol, R. Llera Blanes & R. Sarró (Eds.), *Atlantic perspectives: Spirits, places and heritage* (pp. 172-195). Berghahn Books.

- Dussel, E. (1986). La religiosidad popular latinoamericana (hipótesis fundamentales). *Revista Cristianismo y Sociedad*, 24(88), 103-112.
- Encuesta de Opinión Pública de CID Gallup. (2022). *Evangélicos en Latinoamérica* [Gráfico].
Twitter. <https://x.com/cidgallup/status/1553063419774902272/photo/1>
- Estermann, J. (1998). *Filosofía Andina. Estudio intercultural de la sabiduría autóctona andina*. Abya Yala.
- Estermann, J. (2003). Religión como Chakana. El inclusivismo religioso en los Andes. *Estudios Aymaras*, 2(66), 4-26.
- Flores, G. (2022). RAYMIKUA: Racismos e identidades. En T. Arteaga & A. Japón (Eds.), *RAYMICUNA en los Andes* (pp. 29-40). Universidad de Cuenca.
- Frigerio, A., Juárez, H., & Sampaio, D. (2022). Religiones afrolatinoamericanas: reivindicaciones de etnicidad, espiritualidad y género en la tradición orisha. *Perspectivas Afro*, 2(1), 12-17.
- Gullo Omodeo, M. (1997). Religión política y constitucional en Estados Unidos. *THEMIS Revista de Derecho*, 35, 197-207.
- Instituto Nacional de Estadística y Censos (INEC). (2022). *Filiación religiosa*. <https://www.ecuadorencifras.gob.ec/filiacion-religiosa/>
- Japón, A., Quichimbo, F., & Orellana, A. (2022). La recuperación de los RAYMIKUA desde la educación intercultural. En T. Arteaga & A. Japón (Eds.), *TAWA PACHA RAYMI. Sabidurías y transcendencia* (pp. 67-82). Universidad de Cuenca.
- Marzal, M. (2005). *Los santos y la transformación religiosa del Perú colonial*. Pontificia Universidad Católica del Perú, Comisión de Fe y Cultura.
- Matory, L. (1998). Yorubá: as rotas e as raízes da nação transatlantica, 1830-1950. *Horizontes Antropológicos*, 4(9), 263-292.
- Morocho, P. (2008). El Corpus Christi del pueblo indígena del Cañar. En CIDAP (Ed.), *Artesanías de América* (pp. 101-118). Centro Interamericano de Artesanías y Artes Populares.
- Ramírez, J. (2009). *El campo religioso latinoamericano y caribeño. Efectos de la globalización neoliberal*. Consejo Latinoamericano de Ciencias Sociales (CLACSO).
- Ratzinger, J. (1995). *Verdad, valores, poder*. Ediciones Rialp.
- Ríos, S. (2008). Los elementos andinos y católicos en el contexto de las fiestas religiosas populares. En CIDAP (Ed.), *Artesanías de América* (pp. 179-210). Centro Interamericano de Artesanías y Artes Populares.
- Rosbach de Olmos, L. (2014). Cruces y entrecruzamientos en los caminos de los orichas: tradiciones en conflicto. *Indiana*, 31, 86-107.
- Segato, R. (1993). La religiosidad candomblé la tradición afrobrasileña. *Perfiles Latinoamericanos*, 2(3), 133-164.

- Simbaña, F., & Caral, N. (2022). RAYMIKUA: Racismos e identidades. En T. Arteaga & A. Japón (Eds.), *TAWA PACHA RAYMI. Sabidurías y transcendencia* (pp. 83-98). Universidad de Cuenca.
- Stark, A. (2012). La religión en los orígenes de los Estados Unidos: hacia la democracia de los elegidos. *Pencopolitana de Estudios Históricos y Sociales*, 2, 35-54.
- Trigilio, J., & Brighenti, K. (2007). *Catolicismo para Dummies*. For Dummies.
- Untoja Choque, F., & Mamani Espejo, A. (2000). *Pacha en el pensamiento aymara*. Editorial de los Diputados.
- Vaticano. (2012). *Los principios del catolicismo*. <https://www.vatican.va/content/vatican/es.html>
- Vatican News. (2023, 4 de marzo). Continúa el crecimiento de los católicos en el mundo, 1.378 millones en 2021. *Vatican News*. <https://www.vaticannews.va/es/vaticano/news/2023-03/anuario-pontificio-estadistica-iglesia-mundo-2023-libreria.html>
- Viteri, L. (2012). *La religión "cristiana-evangélica" y su influencia en las comunidades indígenas-campesinas del Ecuador andino* [Tesis de grado, Universidad Andina Simón Bolívar].
- Zapata Olivella, M. (1989). *Las claves mágicas de América*. Biblioteca Nacional de Colombia.

El anarquismo decimonónico frente a la religión Nineteenth-century anarchism confronting religion

Boris Briones Soto

boris.briones@ucn.cl

Universidad Católica del Norte

ORCID ID 0000-0002-2205-8065

Resumen: El anarquismo, como filosofía política y económica, durante el siglo XIX abordó diversos temas vinculados al Estado, la organización social, el sindicalismo y los movimientos obreros. Entre sus preocupaciones centrales estuvo la religión y la idea de Dios, entendidas como estructuras de poder que condicionaban la vida en sociedad. Durante la segunda mitad del siglo XIX, Pierre-Joseph Proudhon y Mijaíl Bakunin desarrollaron una crítica profunda a los sistemas de creencias y a la institucionalidad religiosa, considerando la existencia de una estrecha relación con la dominación política y social. Este trabajo analiza sus principales reflexiones sobre lo divino, explorando cómo estas ideas se insertaron en el desarrollo del anarquismo histórico y en la forma en que los anarquistas decimonónicos entendieron el papel de la religión en un contexto de profundos cambios ideológicos y políticos.

Palabras clave: Anarquismo decimonónico; religión y poder; crítica anarquista; Proudhon y Bakunin; Estado e institucionalidad religiosa; historia del pensamiento político.

Abstract: Anarchism, as a political and economic philosophy, in the 19th century addressed various issues related to the state, social organization, syndicalism, and labor movements. Among its central concerns were religion and the idea of God, understood as structures of power that shaped social life. During the second half of

the 19th century, Pierre-Joseph Proudhon and Mikhail Bakunin developed a profound critique of belief systems and religious institutions, arguing that they were closely linked to political and social domination. This study examines their main reflections on the divine, exploring how these ideas became part of the development of historical anarchism and how 19th-century anarchists conceptualized the role of religion in a context of profound ideological and political transformations.

Keywords: 19th-century anarchism; religion and power; anarchist critique; Proudhon and Bakunin; state and religious institutions; history of political thought.

Introducción

Lo que hoy entendemos por anarquismo ha transitado por distintos cambios históricos, que en ocasiones son difíciles de dilucidar. Por lo mismo, la palabra *anarquía* tiene cierta carga peyorativa, considerando que “(...) se emplea muchas veces cuando se quiere señalar que algo está descontrolado y sin rumbo, anárquico se configura como sinónimo de desbarajuste, de caos” (Paniagua, 2012, p. 16), si a lo anterior sumamos los errores en los que caen ciertos sectores políticos e incluso la prensa cuando hay manifestaciones públicas, el concepto de anarquismo ha sido ampliamente agraviado en la actualidad.

Si los anarquistas asocian el anarquismo como una amplia gama de doctrinas que reprueban las instituciones de gobierno como innecesarias y perjudiciales para el desarrollo económico y social, sus enemigos han equiparado la anarquía con el caos social y el desorden violento. Esta descalificación del anarquismo, así como la categorización de los anarquistas como descerebrados instigadores del caos social, ha tenido una larga historia en Europa Occidental. (Mac Laughlin, 2017, p. 12)

En Europa durante el siglo XIX, nos encontraremos con un auge de propuestas sociales, económicas y políticas que buscan resolver los problemas acarreados después de la Revolución Francesa, lo que implicó importantes cambios (Xu, 2024), particularmente en temas ideológicos. Este momento histórico estuvo marcado por una transformación económica producto de los procesos de industrialización y su masificación, esto llevó a la aparición y auge de los sindicatos, desde donde surgieron diversos intelectuales y también cierto “mesianismo revolucionarista” (Zorrilla, 1988). En todo este contexto surgen dos principales ideas o modelos, en principio cercanas, pero posteriormente distantes: el anarquismo y el marxismo.

Sería demasiado ambicioso en este trabajo querer explicar en detalle los fundamentos y propuestas de cada modelo, teniendo en consideración especialmente que el anarquismo se formaliza en el siglo XIX con un sistema ideológico más claro, no obstante se logran evidenciar ideas protoanarquistas en otros momentos históricos anteriores y en lugares muy diversos del mundo (Cappelletti, 2008). Por lo anterior, en este trabajo solo nos situaremos en analizar el pensamiento decimonónico anarquista en un área particular: la religión y Dios. Al parecer este tema para algunos anarquistas clásicos fue de gran relevancia y no podía estar separado del imaginario político y social.

Contexto Siglo XIX

Europa durante el siglo XIX se vio marcada por lo que fue la Revolución Francesa y la caída de las instituciones del Antiguo Régimen, también por los procesos independentistas de América, que de uno u otro modo, influyeron en los pensamientos políticos y sociales decimonónicos. Había pasado, con gran auge, el siglo de las luces –siglo XVIII–, marcado profundamente por diversas corrientes intelectuales, el uso de la razón para la aplicación del método científico y la difusión de distintos saberes en aristas muy variadas (McLean, 2010; Outram, 2009).

Desde finales del siglo XVIII sucedieron cambios fundamentales, por el declive gradual de las monarquías y los imperios, que dieron paso a democracias y nuevas formas de autoritarismo (Stearns, 2012, p. 132). El siglo XIX es en parte heredero de la *Ilustración*, con ciertos matices y no exento de

problemas, para el historiador Geoffrey Bruun (1964), entre 1830 y 1848 Europa se vio marcada por lo que denomina “avances y frustraciones”. Destacan en este periodo ciertos impulsos humanitarios como la abolición de la esclavitud en Inglaterra en 1833, “con compensaciones de veinte millones de libras a los dueños de esclavos coloniales” (Bruun, 1964, p. 54). Mientras Francia e Inglaterra avanzaban en una reforma electoral y un sufragio amplio, otros lugares de Europa oriental lo consideraban algo peligroso (Bruun, 1964, p. 58).

La Revolución de Julio, situada en París durante 1830, fue fundamental para el posterior desarrollo de las ideas anarquistas y la aparición de los primeros textos de Pierre-Joseph Proudhon. En este período hay una disputa entre absolutistas y liberales (Croce, 1965) que se interesaban en evaluar las formas para llegar al poder, más que por los problemas estructurales.

Durante la primera mitad del siglo XIX, según Hobsbawm (2011) encontraremos tres principales olas revolucionarias en occidente. La primera situada entre 1820 y 1824 en el mediterráneo europeo y en la América española, la segunda situada entre 1829 y 1834 en Europa al oeste de Rusia y en Norteamérica. Por último, en 1848 se produce lo que el autor considera “la mayor de las olas revolucionarias” (2011, p. 119), la cual se desató casi de forma paralela en Francia, en gran parte de Italia, en los Estados alemanes, en fracción del Imperio de los Habsburgo y en Suiza.

En este clima de constantes cambios, del nacimiento de ideas disímiles, aparece lo que hoy denominaremos *anarquismo histórico*, para diferenciarlo de los procesos posteriores y actuales. Por supuesto, en estos climas revolucionarios existió un fuerte deseo de liberación, que estuvo acompañado por la constante búsqueda para identificar a los opresores. Es en ese contexto que aparece la religión como uno de ellos. Este trabajo analiza las obras de dos autores en particular: Mijaíl Bakunin y Pierre-Joseph Proudhon, anarquistas teóricos por excelencia, considerados parte del anarquismo moderno.

Proudhon

Considerado el primer teórico del anarquismo moderno (Paniagua, 2012), publicó en 1840 su obra fundamental *Qu'est-ce que la propriété? ou Recherche sur le principe du Droit et du Gouvernement*, texto que lo llevará a definirse directamente como anarquista, dejando frases célebres como la reflexión introductoria:

Si tuviera que contestar a la siguiente pregunta: *¿Qué es la esclavitud?*, y respondiese en pocas palabras *es el asesinato*, mi pensamiento sería comprendido de inmediato. No necesitaría, ciertamente, grandes razonamientos para demostrar que la facultad de quitar al hombre el pensamiento, la voluntad, la personalidad, es un derecho de vida y muerte, y que hacer esclavo a un hombre es asesinarlo. ¿Por qué razón, sin embargo, no puedo contestar a la pregunta *¿Qué es la propiedad?*, diciendo concretamente es el robo, sin tener la certeza de no ser comprendido, aun cuando esta segunda respuesta no sea más que una simple transformación de la primera? (Proudhon, 1975, p. 29)

Para Mac Laughlin “el famoso aserto de Proudhon, —la propiedad es un robo—, tiene que tomarse en su contexto histórico y hasta regional. La propiedad privada de la tierra sólo había aparecido en Francia en la década de 1760” (2017, p. 59).

Este es solo el contexto inicial de un extenso trabajo que discute el principio del gobierno, las leyes naturales y la fundamentación —filosófica si se quiere— de la propiedad en general. No es discusión de este artículo el valor de la obra para el pensamiento anarquista posterior, que sin duda tendrá un apoyo importante en quien es considerado el primer teórico del anarquismo moderno.

En la primera parte de este libro, Proudhon señala que uno de los principios antiguos, que resulta muy difícil de superar es la religión y la noción de Dios. Comienza discutiendo:

Todos los hombres creen en Dios; este dogma corresponde a la vez a la conciencia y a la razón. Dios es para la humanidad un hecho tan primitivo, una idea tan fatal, un principio tan necesario como para nuestro entendimiento lo son las ideas categóricas de causa, de substancia, de tiempo y de espacio. A Dios nos lo muestra nuestra propia conciencia con anterioridad a toda inducción del entendimiento, de igual modo que el testimonio de los sentidos nos prueba la existencia del Sol anticipándose a todos los razonamientos de la física. La observación y la experiencia nos descubren los fenómenos y sus leyes; el sentido interno sólo nos revela el hecho de su existencia. La humanidad cree que Dios existe, pero ¿qué es lo que cree al creer en Dios? En una palabra: ¿qué es Dios? (1975, p. 37)

Resultaba difícil imaginar que en un libro que trata sobre cuestiones relativas a la propiedad y el derecho en general, se pudiera encontrar una reflexión de tal profundidad, cuando la idea de Dios no es el tema central de este texto, pero como señalamos, se encontrará presente en gran parte de la obra anarquista. Este cuestionamiento, que ha sido ampliamente discutido por la historia y la filosofía (Casales & Bastos, 2022), no parece tener una respuesta sencilla. Sin embargo, estas ideas nos aproximan a un pensamiento crítico con el legado religioso en la era posrevolucionaria, siendo también un evidente legado del método científico y la obra de Descartes.

Continuando con sus exposiciones, Proudhon argumenta:

La noción de la divinidad, noción primitiva, unánime, innata en nuestra especie, no está determinada todavía por la razón humana. A cada paso que avanzamos en el conocimiento de la naturaleza y de las causas, la idea de Dios se magnifica y eleva; cuanto más progresa la ciencia del hombre, más grande y más alejado le parece Dios. El antropomorfismo y la idolatría fueron consecuencia necesaria de la juventud de las inteligencias, una teología de niños y de poetas. Error inocente, si no se hubiese querido hacer de él una norma obligatoria de conducta, en vez de respetar la libertad de creencias. (1975, p. 37)

En esta primera parte de la exposición, el autor sorprende al presentar la propuesta de que la religión es inherente al ser humano, casi de manera natural. Ideas que han sido discutidas por la historia de las religiones en la prehistoria (Wunn, 2012). Proudhon propone que esta especie de estadio primitivo ha ido cambiando producto de la evolución de las sociedades, postulados que parecen tomados del trabajo evolucionista de Lewis Morgan y los tres estadios evolutivos de salvajismo, barbarie y civilización, ampliamente debatidos por la antropología y etnología. El problema y la novedad a la vez, radica en que Proudhon publica este texto en 1840 y la obra de Morgan no verá la luz hasta 1877. Es curioso que el pensador anarquista presente una visión algo evolucionista, incluso considerando que *On the Origin of Species* de Darwin no aparecería hasta 1859. Al parecer, sus críticas son todavía más profundas:

Pero el hombre, después de haber creado un Dios a su imagen, quiso además apropiárselo; no contento con desfigurar al Ser Supremo, le trató como su patrimonio, su bien, su cosa; Dios, representado bajo formas monstruosas, vino a ser en todas partes propiedad del hombre y del Estado. Éste fue el origen de la corrupción de las costumbres por la religión, y la fuente de los odios religiosos y de las guerras sagradas. Al fin, hemos sabido respetar las creencias de cada uno y buscar la regla de las costumbres fuera de todo culto religioso; esperamos sabiamente para determinar la naturaleza y los atributos de Dios, los dogmas de la teología, el destino del alma, etc., que la ciencia nos diga lo que debemos olvidar y lo que debemos creer. (Proudhon, 1975, p. 37)

Este punto merece una mayor detención, ya que se introduce un concepto propio del siglo XX, el de “ser supremo”, no obstante esta palabra ya había estado resonando durante el Terror en la Revolución francesa con el *culte de l'Être suprême*.

Esta noción es estudiada por la fenomenología en la obra de Van der Leeuw (1931) y particularmente por la escuela romana de historia de las religiones con Pettazzoni (1974 y 1996). Es erróneo argumentar que el ser

supremo es equivalente a Dios, ya que sus características son diferentes, no necesariamente hablamos de un ente masculino, ni necesariamente con rasgos de ser celestial. Más bien parece una crítica a lo institucionalizado, lo que se evidencia cuando menciona “las guerras sagradas”, parte del imaginario legado por el medioevo.

Proudhon profundiza su crítica al entrar en los conceptos de *pecado* y de *culpa*. Aborda también la idea de que en la historia de la humanidad se encuentra una necesidad latente del mal, ideas que parecen emanadas de una teodicea:

El hombre, dicen los teólogos, ha pecado en su origen; su raza es culpable de una antigua prevaricación. Por esa falta, la humanidad ha degenerado; el error y la ignorancia han llegado a ser sus inevitables frutos. Leyendo la historia, encontraréis en todos los tiempos la prueba de esta necesidad del mal en la permanente miseria de las naciones. El hombre sufre y sufrirá siempre; su enfermedad es hereditaria y constitucional. Usad paliativos, emplead emolientes; no hay remedio eficaz. (1975, p. 38)

En este pequeño fragmento se puede apreciar el sentido de la obra del autor anarquista, donde su crítica generalizada a la religión está basada en el cristianismo por excelencia y se desprende de la misma idea con la que aborda el concepto de Dios. Continúa en la misma línea esbozando otra crítica: “[...] este razonamiento no sólo es propio de los teólogos; se encuentra en términos semejantes en los escritos de los filósofos materialistas, partidarios de una indefinida perfectibilidad” (1975, p. 38). Será esta una de las primeras críticas indirectas a los trabajos de Marx, que como sabemos, se profundizarán con el tiempo, desarrollando una serie de disputas entre ambos, muy polémicas en su momento.

La exposición Proudhon avanza afirmando que las consecuencias del pecado de Adán son transmitidas a su descendencia, señala como principal problema la ignorancia (1975, p. 38). Reflexionará diciendo “¿Qué es la justicia? Los teólogos contestan: Toda justicia viene de Dios. Esto es cierto, pero nada enseña” (1975, p. 39). Si bien los cuestionamientos de Proudhon

parecen avasalladores y a la vez complejos, reflejan una profundidad de reflexión y pensamiento propio de su contexto, que incluso va más allá del anticlericalismo europeo del momento. Se entiende como un preámbulo esencial para abordar el fondo de sus postulados, al parecer no es posible separar al Estado de la religión, ni es posible separar al individuo de sus creencias, como si los dogmas coartaran el pensamiento y sumieran a la persona en una especie de miseria irracional.

También llegará a referirse a la figura principal del cristianismo, que denominará Palabra de Dios, de quien dirá, era un hombre que:

predicaba por todas partes que la sociedad expiraba, que el mundo iba a transformarse; que los maestros eran falaces, los jurisconsultos, ignorantes, los filósofos, hipócritas y embusteros; que el amo y el esclavo eran iguales, que la usura y cuanto se le asemejaba era un robo, que los propietarios y concupiscentes serían alcanzados por el fuego eterno, mientras los pobres de espíritu y los virtuosos habitarían en un lugar de descanso. (Proudhon, 1975, p. 41)

Es evidente que el discurso del autor se enlaza con las enseñanzas de Jesús en el Nuevo Testamento, para que tenga cierto sentido con el pensamiento ácrata. Sobre el Jesús histórico se ha dicho mucho en historia de las religiones, también se deben considerar los trabajos que han relacionado el cristianismo con el anarquismo (Boehrer, 2001; Ellul, 2011; Tarleton, 1993), trabajos que por supuesto han tenido en consideración una lectura crítica y atenta de los evangelios.

La exposición de Proudhon continúa indicando que Jesús fue denunciado y preso como “enemigo público por los sacerdotes y los doctores de la ley, quienes tuvieron la habilidad de hacer que el pueblo pidiese su muerte” (p. 41). Esto evidencia el conocimiento y lectura que el autor había tenido del Nuevo Testamento, pero además una revisión detallada de estos textos. En esta última frase se esboza el pensamiento anarquista: advertir que alguien se ha convertido en enemigo de la autoridad y cómo esa misma autoridad manipula al pueblo. Es la forma en la que los anarquistas entienden que las personas pierden la identidad.

Luego de una amplia reflexión sobre este tema, presenta una crítica a la conformación del cristianismo primitivo, indicando que caería en una idolatría y una mitología semi-poética, llena de discordias. “En vez de atenerse a las consecuencias prácticas de los principios de moral y de autoridad que *Palabra de Dios* había proclamado, se distrajo el ánimo en especulaciones sobre su nacimiento, su origen, su persona y sus actos” (Proudhon, 1975, p. 42). Posterior a una severa crítica a los estudios sobre los textos neotestamentarios, cierra con una potente aseveración: “nació la teología, que se puede definir como la ciencia de lo infinitamente absurdo” (Proudhon, 1975, p. 42).

Es curioso que la primera obra de Proudhon, que sentó las bases del anarquismo histórico europeo, tenga un apartado tan grande dedicado a la religión. No ha de extrañarnos entonces que el posterior desarrollo del pensamiento ácrata sea en base a un ateísmo científico. Por supuesto no se detiene aquí, vendrá la crítica más profunda al catolicismo en ¿Qué es la propiedad? (1975).

La verdad cristiana no va más allá de la edad de los apóstoles; el Evangelio, comentado y simbolizado por los griegos y latinos, cargado de fábulas paganas, llegó a ser, tomado al pie de la letra, un conjunto de contradicciones; y hasta la fecha el reino de la *Iglesia infalible* ha sido el de las tinieblas. Dícese que las puertas del *infierno* no prevalecerán, que la *Palabra de Dios* se oirá nuevamente, y que, por fin, los hombres conocerán la verdad y la justicia; pero, para entonces, habrá terminado el catolicismo griego y romano, de igual modo que a la luz de la ciencia desaparecen los fantasmas de la opinión.

Los monstruos que los sucesores de los apóstoles tuvieron por misión exterminar, repuestos de su derrota, reaparecieron poco a poco, merced al fanatismo imbécil y a la deliberada connivencia de los clérigos y de los teólogos. (p. 42)

Este mensaje va dirigido al cristianismo, ya no es una referencia vaga e inexacta de *religión* o de Dios. Hay varias críticas que se desprenden de este fragmento. En primer lugar la carga simbólica del Nuevo Testamento y el

problema de los añadidos de otros ciclos míticos previos, cuestión que fue estudiada ampliamente por los detractores de la historicidad de Jesús durante el siglo XX (Piñero, 2018, 2020). Se esboza de igual modo una profunda crítica a la carga grecorromana que tiene el cristianismo, lo que, según el autor, una vez masificada la verdad y la justicia, podría llevar al fin de esta religión. La crítica final, parece un rechazo a lo institucionalizado, también a lo ocurrido en el auge del cristianismo medieval, época de persecuciones y problemas derivados del establecimiento de una religión oficial.

Posterior a esta exposición, Proudhon esboza los acontecimientos que llevan a la Revolución Francesa, criticando fuertemente los privilegios reales y la tiranía, además de lo que denomina “intolerancia sacerdotal” (1975, p. 42).

Proudhon no sólo escribirá respecto a la propiedad y el sentido del derecho. Otra de sus obras fundamentales se titula *Escritos federalistas*, ideas redactadas a principios de la década de 1860 y que ven la luz en una edición crítica en castellano el 2011 a cargo de Jorge Cagiao y Conde. En este trabajo, el pensador anarquista aborda temas como las formas de federación y de unidad, pasando por la política unitaria y la prensa unitaria. Como era de esperarse, hay un apartado dedicado al factor religioso, titulado *El Papado y las Sectas Religiosas*, aborda aquí la problemática de la Iglesia en Italia y manifiesta su preocupación por las estrechas relaciones que pueden tejer los líderes políticos y religiosos, hace hincapié en que la libertad de cultos de un Estado, abre la puerta para que un sacerdote en su calidad de ciudadano, pueda ser elegido como representante del pueblo (Proudhon, 2011, p. 326). Esto por supuesto, preocupa al autor, ya que sería entrar en el terreno de la laicidad —y el laicismo—. Frente a esto señala que:

la federación es libertad por excelencia, pluralidad, división, gobierno de uno mismo por sí mismo. Su máxima es el derecho, no otorgado por la Iglesia, intérprete del cielo, ni definido por el príncipe, representante de la divinidad y brazo del Santo Padre, sino determinado por el libre contrato. En este sistema, la ley, el derecho, la justicia son el estatuto arbitral de las voluntades, estatuto superior,

en consecuencia, a toda autoridad y creencia, a toda Iglesia y religión, a toda unidad, puesto que la autoridad y la fe, la religión y la Iglesia, siendo patrimonios exclusivos de la conciencia individual, se sitúan por esta misma razón por debajo del pacto, expresión de consentimiento universal, la más alta autoridad entre los hombres. (2011, p. 330)

Al parecer, solo en la propuesta federativa está la respuesta para la aplicación de un Estado de Derecho acabado e impoluto. Es en estas condiciones donde se alcanzará una convivencia colectiva libre. Estos principios federalistas influenciaron el pensamiento político en distintos lugares del mundo, que buscaron mediante ensayos constitucionales, abordar la cuestión de la organización interna de una nación. Por supuesto que la preocupación por lo religioso sigue presente en la madurez intelectual de Proudhon, de ahí a que señale en el contexto de su análisis de la relación Iglesia-Estado: “el sufragio universal no hará nunca de una república federativa un Estado pontificio” (2011, p. 331). Al parecer, la preocupación central del anarquista se aclara con esta sencilla afirmación. Sus inquietudes no se relacionan con la religión entendida como un sistema de creencias y cuestiones espirituales, sino más bien con la autoridad del catolicismo, los que se encuentran en la cabeza del poder en la institucionalidad cristiana, en este caso particular, el papado. Aquí viene una de sus propuestas más interesantes, Proudhon señala que la religión no es necesaria para la moral, suponiendo que el Evangelio no ha dicho la última palabra del derecho, plantea que una justicia basada en la adoración, es una justicia inexacta (2011, p. 331).

Bakunin

Considerado un promotor, impulsor y agitador inagotable, un revolucionario sin igual (Mintz, 2006, p. 8). Es sencillo ubicar a Bakunin en el anarquismo, pero también tiene un espacio importante en el *socialismo libertario* (McLaughlin, 2002). Son dos las obras de Bakunin que nos interesan para efectos de este trabajo. En primer lugar *Fédéralisme, socialisme, antithéologisme* publicada en 1867, y su obra más destacada en el ámbito de lo religioso *Dieu et l'État* de publicación póstuma en 1882. Bakunin, durante sus estancias en

París compartió con Marx, Engels y por supuesto con Proudhon, con quien tendrá más afinidad intelectual, McLaughlin llega a señalar que el intelectual ruso puede ser considerado un proudhoniano de izquierda (2002, p. 72).

Uno de sus principales aportes conceptuales es lo que denominara *antiteologismo*, Bakunin inicia su reflexión señalando:

No siendo todas las religiones con sus dioses más que una creación de la fantasía mística y crédula del hombre, que no ha llegado todavía a la altura de la reflexión pura y del pensamiento libre apoyado en la ciencia, el cielo religioso es un milagro en el que el hombre, exaltado por la fe, halló largo tiempo su propia imagen, pero agrandada y trastornada, es decir divinizada.

La historia de las religiones, la de la grandeza y decadencia de los dioses que se han sucedido, no es, pues, otra cosa que la historia del desenvolvimiento de la inteligencia y de la conciencia colectiva de los hombres. A medida que descubrían, sea en ellos, sea fuera de sí, una fuerza, una capacidad, una cualidad cualquiera la atribuían a sus dioses, después de haberla engrandecido, ampliado sobre toda medida, como hacen ordinariamente los niños, por un acto de fantasía religiosa. (1979, p. 79)

Desde un primer momento queda claro que el autor entiende la religión como algo negativo para el ser humano, como una especie de lastre que se puede superar. Una concepción cargada de ideas preconcebidas, que generalizan *religión* y que no tienen sustento más que en unos pocos ejemplos que el autor irá desentrañando. Parte del supuesto, al igual que Proudhon, de que todos los seres humanos tienen una tendencia a la religión, en cuanto creencia natural, por el simple hecho de ser *Homo sapiens*. Por supuesto que en los tiempos que Bakunin escribe, no hay avances en arqueología cognitiva ni en estudios sobre religión en la prehistoria, motivo por el cual hoy su afirmación nos parecería inexacta. Continúa con la errónea idea de que toda religión es teísta, en tiempos en que existían y eran conocidas grandes tradiciones no teístas como el budismo, pero al parecer su análisis considera solo el cristianismo, entendida por el autor como:

la religión por excelencia, porque expone y manifiesta la naturaleza y la esencia de toda religión, que son: el empobrecimiento, el aniquilamiento, el sometimiento sistemáticos, absolutos, de la humanidad en beneficio de la divinidad, principio supremo, no ya de toda religión, sino también de toda metafísica teísta o panteísta. (1979, p. 80)

Es indiscutible que el autor conoce principios filosóficos contundentes, por esto entiende la religión como alienación, como algo que coarta la libertad humana, que por supuesto sería contraria a la revolución, lo que clarifica al señalar que “la Historia nos demuestra que los sacerdotes de todas las religiones, menos los de las iglesias perseguidas, han sido los aliados de la tiranía” (1979, p. 81). Ya en el siglo XIX se podía vislumbrar esto con claridad con algunos ejemplos históricos.

Al igual que en Proudhon, el problema se relaciona con el poder, para Bakunin el cristianismo es una religión fundada y bautizada en sangre más que cualquier otra, “que se cuenten los millones de víctimas que esta religión del amor y del perdón ha inmolado a la venganza cruel de su Dios” (1979, p. 82).

No obstante, Bakunin va más allá e interroga sobre una crítica a los cimientos religiosos:

Mientras no podamos darnos cuenta del modo como se produjo la idea de un mundo sobrenatural o divino, y como debió de producirse necesariamente en el desenvolvimiento natural del espíritu humano y de la humana sociedad por la Historia, podremos estar científicamente convencidos de lo absurdo de esa idea, pero no lograremos destruirla nunca en la opinión del mundo, porque sin ese conocimiento no conseguiríamos atacarla nunca en las profundidades mismas del ser humano. (1979, p. 93)

Este fragmento nos ayuda a entender lo que ya venía señalando el autor, la idea de persistencia de la religión en la humanidad, entendida como un producto del desarrollo histórico. Si bien considera y acepta que en su

presente es imposible superar dicha idea, deja entrever que más adelante las cosas podrían cambiar. Esta postura que parece simplista, es muy profunda, porque la crítica planteada esgrime que no hay que limitarse a la crítica superficial de la religión, sino que se debe abordar su origen histórico, la historicidad y las estructuras sociales que permitieron en el pasado su manifestación y existencia. Al parecer estos elementos podrían ser también psicológicos.

Su indagatoria por el origen de la religión, es evidentemente más profunda que el trabajo de Proudhon, esto se manifiesta en una dialéctica constante con la historia:

A juzgar por los relatos unánimes de los viajeros que, desde el siglo pasado, han visitado las islas de Oceanía, como de los que en nuestros días han penetrado en el interior de África, el fetichismo debe de ser la primera religión, la de todos los pueblos salvajes que se han alejado menos del estado natural. Pero el fetichismo no es otra cosa que la religión del miedo. Es la primera expresión humana de esa sensación de dependencia absoluta, mezclada al terror instintivo, que hallamos en el fondo de toda vida animal. (Bakunin, 1979, p. 102)

Bakunin interpreta el fenómeno religioso desde una mirada materialista y evolucionista, es evidente que se influencia de la antropología y etnología del siglo XIX, en los estudios de las llamadas religiones étnicas (Briones, 2024, p. 7). Esta postura del autor se alinea con las ideas de Tylor y Frazer. En el fragmento anterior, destaca la frase *religión del miedo*, al parecer encausa el fetichismo no como un proceso racional en las etapas de comprensión del mundo, sino más bien como una reacción emocional ante lo desconocido. Más adelante esta idea se complementa con lo que denomina *reflexión imaginativa del hombre*, explicando que se adoran instintivamente los grandes fenómenos de la naturaleza que ejercen una influencia inmediata en la existencia (Bakunin, 1979, p.109). Bakunin sostiene que el fetichismo se transforma en un –culto de los brujos–, lo que interpretaremos como chamanismo. Aquí hay un paralelismo interesante, ya que dice que hoy – 1867– priman los espiritistas, los médiums y los clarividentes (Bakunin, 1979,

p. 109). Recordemos que el siglo XIX en Francia fue el epicentro de la doctrina espírita con su exponente Allan Kardec, momentos de masificación de estos movimientos por Europa, que llenaban teatros haciendo levitar mesas.

Para Bakunin, en –la brujería primitiva–, la divinidad es inseparable de la persona que ejerce como brujo, “él mismo era dios como el fetiche” (1979, p. 110), esto desencadena una búsqueda de Dios. Señala que con el paso del tiempo, esta constante lleva a la abstracción y a una autognosis. Mediante estos procesos dice, “es como ha debido de nacer el panteísmo religioso de los hindúes” (1979, p. 111). Es evidente que Bakunin, conoce variados sistemas religiosos y que su análisis es más complejo que el de Proudhon en toda regla. Continúa reflexionando con la idea de que en la evolución del pensamiento religioso, desde el fetichismo hasta los fenómenos naturales y astronómicos, nos lleva a generar la idea de un dios espiritual, extraterreno e invisible.

Ahora presenta, por primera vez una divinidad universal: el ser de los seres, sustancia y creador de todos esos seres restringidos y particulares, el alma universal de todo el universo, el Gran Todo. He aquí, pues, el verdadero Dios que comienza y con él la verdadera religión. (1979, p. 111)

Aquí, mediante lo que denomina un análisis histórico, se propone buscar la verdadera naturaleza de la divinidad. No lo logra, pero llega a la conclusión de que “Dios es la abstracción absoluta, es el propio producto del pensamiento humano (...) el ser único y supremo” (Bakunin, 1979, p. 113). Más adelante dirá “Una vez aparecido Dios, el desenvolvimiento sucesivo y progresivo de las diferentes teologías se explica naturalmente como el reflejo del desenvolvimiento de la humanidad en la Historia” (Bakunin, 1979, p. 119). Es esta idea central la que reflejará su obra consagratoria en materia religiosa: *Dios y el Estado* (1882). Muchas ideas se repiten, complementan o profundizan. Esto tiene que ver con que Bakunin escribió de manera interrumpida en muchas ocasiones y que, gran parte de sus textos quedaron en manuscritos y solo vieron la luz cuando su viuda entregó una serie de papeles para que se

publicaran las obras que hoy tenemos disponibles, con títulos no necesariamente exactos (Maíz, 2021).

En *Dios y el Estado*, Bakunin comienza con una reflexión crítica de la Biblia, indicando que “es un libro muy interesante y a veces muy profundo cuando se lo considera como una de las más antiguas manifestaciones de la sabiduría y de la fantasía humanas” (1992, p. 34). Para el anarquista, es evidente que el problema de la creencia está relacionada con el desconocimiento, así lo evidencia cuando dice que la idea de un Dios creador se conserva mayormente en las poblaciones rurales, ya que “el pueblo, desgraciadamente, es todavía muy ignorante; y es mantenido en su ignorancia por los esfuerzos sistemáticos de todos los gobiernos, que consideran esa ignorancia, no sin razón, como una de las condiciones más esenciales de su propia potencia” (Bakunin, 1992, p. 42). Esto lo podemos interpretar en tres niveles. En primer lugar, la relación entre religión y educación; para Bakunin, la creencia en un Dios se sustenta por la falta de acceso al conocimiento, una concepción materialista, ya que entiende la religión no como resultado de una verdad revelada ni una especie de necesidad metafísica, sino un producto de la ignorancia. En segundo lugar, encontramos el papel de los gobiernos (tema central en el anarquismo), ya que existe un esfuerzo deliberado por mantener esa ignorancia, entonces la religión es necesaria para el poder. La autoridad religiosa está estrechamente relacionada con la autoridad política y ambas se refuerzan. El Estado no solo tolera la religión, sino que la promueve. Por último, la función de la religión en la estructura de poder. Para Bakunin la religión no parece ser concebida como un fenómeno cultural o teológico, más bien una herramienta funcional al dominio, una especie de sumisión. Esta tesis se anticipa a las propuestas de Foucault. Es evidente la lectura de las escrituras que Bakunin tuvo durante su vida, considerando esta obediencia como se entiende en Romanos 13:1-2.¹ Los cuestionamientos se entienden al comprender que el autor abogaba por construir una sociedad donde se excluyera a la religión y a Dios de los asuntos públicos (Kneper, 2015, p. 408). “¿Cómo se transita desde la crítica materialista de la idea de Dios hacia la

¹ "Que toda persona se someta a las autoridades que [nos] presiden, pues no hay autoridad a no ser bajo la acción de Dios, y las que existen han sido constituidas por Dios". (Cantera Iglesias, 2009)

crítica del poder estatal? Bakunin traza una vía de argumentación centrada en la relación asimétrica y abusiva del individuo con su Dios” (Jiménez, 2021, pp. 497-498).

Hay otra razón que explica y que legitima en cierto modo las creencias absurdas del pueblo. Es la situación miserable a que se encuentra fatalmente condenado por la organización económica de la sociedad en los países más civilizados de Europa. Reducido, tanto intelectual y moralmente como en su condición material al mínimo de una existencia humana, encerrado en su vida como un prisionero en su prisión, sin horizontes, sin salida, sin porvenir mismo, si se cree a los economistas, el pueblo debería tener el alma singularmente estrecha y el instinto achatado de los burgueses para no experimentar la necesidad de salir de ese estado; pero para eso no hay más que tres medios, dos de ellos ilusorios y el tercero real. Los dos primeros son el burdel y la iglesia, el libertinaje del cuerpo y el libertinaje del alma; el tercero es la revolución social. (Bakunin, 1992, p. 43)

Este fragmento es probablemente uno de los más conocidos del autor y, a la vez, uno de los más radicales. Se relaciona la religión con la miseria social, entendida como refugio de los oprimidos. Algo que se relaciona con la idea de Marx respecto a la religión como el suspiro de la criatura oprimida, el opio del pueblo (Bowker, 2006). ¿Durante los períodos de crisis económicas y sociales hay un resurgimiento religioso? Es una pregunta interesante que debe ser contestada en una investigación particular, al parecer Bakunin así lo interpreta. Las personas buscan un escape a la realidad, una especie de sistema hedonista que permite alcohol y una sexualidad sin control: placer inmediato. También está la opción de la sumisión espiritual, lo que denomina “libertinaje del alma”, un escape psicológico, ilusorio, intangible. Frente a estas dos ideas, la salida viable es la revolución social. La problemática que aqueja a las personas es la miseria, la pobreza, donde un grupo particular concentra el poder y la riqueza ¿pero si las personas se tornan conscientes y se organizan? Al parecer es algo que se debe superar, pero no es tan fácil, “en tanto persista la raíz de todos los absurdos que atormentan al

mundo, la creencia en Dios permanecerá intacta, no cesará de echar nuevos retoños” (Bakunin, 1992, p. 51).

Más adelante, en *Dios y el Estado* se replican las ideas de *Federalismo, socialismo, y antiteologismo* (1868), hasta llegar a un interesante apartado titulado “Nota sobre Rousseau”. En este punto el autor critica a los liberales, señalando que nacen de la burguesía y que rinden culto al Estado. “La libertad individual no es, según ellos, una creación, un producto histórico de la sociedad. Pretenden que es anterior a toda sociedad, y que todo hombre la trae al nacer, con su alma inmortal, como un don divino” (Bakunin, 1992, p. 139). La valoración negativa hacia el liberalismo es evidente, si esta libertad fuera inherente al ser humano, ¿cómo se justifica la existencia del Estado como garante y protector de las libertades preexistentes? Es una duda manifiesta incluso para quienes no son anarquistas.²

Bakunin vuelve en su acometida contra la teología:

La teología, que tiene el privilegio de ser absurda, y que cree en las cosas precisamente porque son absurdas, ha puesto por encima de las almas humanas inmortales y por consecuencia infinitas, la infinitud superior, absoluta de Dios. Pero para corregirse, ha creado la ficción de Satanás, que representa precisamente la rebelión de un ser infinito contra la existencia de una infinitud absoluta, contra Dios. Y lo mismo que Satanás se ha rebelado contra la infinitud superior de Dios, los santos anacoretas del cristianismo, demasiado humildes para rebelarse contra Dios, se han rebelado contra la infinitud igual de los hombres, contra la sociedad. (1992, p. 142)

Como si la teología y su estudio fuera un absurdo, irracional y contradictorio. Nuevamente la cuestión de fondo son las estructuras de poder. Aparece el dualismo, el bien y el mal encarnados. La figura de *Satanás*

² Hay una serie importante de acontecimientos durante el siglo XIX que forjaran la relación entre liberalismo y religión. Cf. Martinic, Z. (2014). Liberalismo e Iglesia en Europa durante el siglo XIX. *Tradición y Saber*, 81–99.

como una representación de la rebelión a la autoridad divina. Esta narrativa para Bakunin invisibiliza la posibilidad de una rebelión social legítima: cualquier intento de resistencia se percibe como una revuelta contra el orden cósmico divino, es por tanto una rebelión contra Dios. La figura del cristianismo primitivo en su relato, debe ser entendida como una suerte de rebelión frustrada en colectividad. Tema que ha sido abordado con una renovada perspectiva en los estudios de Fernando Bermejo Barrera (2023) al considerar la figura revolucionaria que presentaron los seguidores de Jesús.

Todos estos cuestionamientos no deben extrañarnos si para Bakunin el Estado moderno tiene un origen teológico (D'Angelo, 2019, p. 128). Como alternativa al sistema religioso, Bakunin proponía la ciencia positiva (Kneper, 2015, p. 408).

Antes de cerrar, el autor es certero al señalar la dicotomía obvia en su realidad europea decimonónica, respecto a las riquezas materiales, planteando que esas riquezas deben ser objeto de desprecio, porque Jesús dijo que no se debían acumular tesoros en esta tierra, porque los tesoros están en el corazón, aquí cierra citando Mateo 19:23-24.

Me imagino la cara que deben poner los piadosos y ricos burgueses protestantes de Inglaterra y de Estados Unidos, de Alemania, de Suiza, al leer esas sentencias tan decisivas y tan desagradables para ellos.

Jesucristo tiene razón; entre la codicia de las riquezas materiales y la salvación de las almas inmortales, hay una incompatibilidad absoluta (...) ¿no vale más renunciar al confort y al lujo que da la sociedad y vivir de raíces, como hicieron los anacoretas, salvando su alma para la eternidad, que perderla al precio de algunas decenas de años de goces materiales? (1992, p. 143)

El texto se interrumpe un poco más adelante, ya que es un manuscrito inacabado de Bakunin. Es así como el anarquista, aquejado por una serie de problemas de salud y en la pobreza, falleció en el hospital de Berna el 1 de

julio de 1876. Hoy sobre su lápida en Suiza descansa la frase “Quien no se atreve con lo imposible, nunca alcanzará lo posible”.

Conclusiones

El anarquismo decimonónico, en su enfrentamiento con la religión y la idea de Dios, reflejó una profunda crítica no solo a la fe en sí misma, sino a su instrumentalización como herramienta de dominación. En este trabajo, hemos explorado cómo Pierre-Joseph Proudhon y Mijaíl Bakunin, dos de los principales exponentes del anarquismo histórico, abordaron el problema de la religión desde sus propias perspectivas filosóficas y políticas.

El problema principal en la obra de Proudhon radica en utilizar el concepto de religión de manera indiscriminada, queriendo agrupar todas las expresiones religiosas en una sola entidad homogénea, que reduce la totalidad del fenómeno religioso al cristianismo y, en particular, al catolicismo. Lo mismo ocurre con su concepción de Dios, que es entendida exclusivamente desde los parámetros de la tradición judeocristiana y asociada con el poder institucionalizado. Su crítica, más que una negación de lo sagrado en sí, es un rechazo a la influencia de la religión en la esfera pública y a su papel en la consolidación de estructuras de dominación. No es casual que Proudhon, heredero de la Revolución Francesa, con un marco cultural forjado en la lucha contra la monarquía y la Iglesia, conciba a la religión principalmente como un mecanismo de control social antes que como un fenómeno cultural o espiritual diverso.

Por su parte, Bakunin lleva la crítica un paso más allá, al concebir la religión como una forma de alienación absoluta, una construcción humana destinada a justificar la sumisión y la desigualdad. Su concepción materialista e histórica de la religión lo lleva a afirmar que la creencia en Dios es un producto del miedo, de la ignorancia y de las condiciones de opresión bajo las cuales ha vivido la humanidad. Para él, la única forma de verdadera emancipación es la superación total de la idea de Dios y la erradicación de la religión como institución, pues la considera inseparable de la jerarquía y el poder político. No obstante, su análisis padece de un problema similar al de Proudhon: una generalización del fenómeno religioso, que ignora la

pluralidad de creencias y prácticas, reduciéndolas a una visión eurocéntrica y religiocéntrica marcada por el cristianismo y su relación con el Estado.

A lo largo del siglo XIX, el pensamiento anarquista mantuvo una tensión constante con la religión, considerándola un obstáculo para la libertad y un soporte de la opresión. Sin embargo, la persistencia de lo religioso en las sociedades modernas plantea preguntas que ni Proudhon ni Bakunin llegaron a responder plenamente. ¿Por qué, a pesar de la secularización progresiva, la religión sigue teniendo un papel central en la vida de muchas personas? ¿Es posible una concepción de lo sagrado que no implique dominio ni sumisión o que se separe completamente de lo político/estatal?

Si bien el anarquismo histórico promovió una postura atea y anticlerical, en tiempos más recientes han surgido reinterpretaciones que dialogan de manera más matizada con el fenómeno religioso. Desde el anarquismo cristiano, que reivindica la figura de Jesús como un revolucionario contrario a la autoridad y la jerarquía, hasta movimientos libertarios que exploran espiritualidades no teístas o alejadas de las instituciones tradicionales, la relación entre anarquismo y religión sigue siendo un campo de debate abierto. El rechazo de los anarquistas decimonónicos a la religión fue, más que un ataque a las creencias individuales, una oposición a la religión como estructura de poder.

Agradecimientos

Este trabajo es el resultado de un proyecto titulado «Conceptos e ideas de Dios y religión según el anarquismo histórico» financiado mediante una beca C de ICALE, Stipendienwerk Lateinamerika-Deutschland e.V.

Referencias

- Bakunin, M. (1979). *Federalismo, socialismo y antiteologismo*. La Piqueta.
- Bakunin, M. (1992). *Dios y el Estado*. Ediciones Júcar.
- Bermejo Barrera, F. (2023). *La invención de Jesús de Nazaret. Historia, ficción, historiografía*. Akal.
- Boehrer, F. (2001). *Christian anarchism and the Catholic Worker movement: Roman Catholic authority and identity in the United States*. Syracuse University.
- Bowker, J. (2006). *Diccionario abreviado Oxford de las Religiones del Mundo*. Paidós.
- Briones Soto, B. (2024). Historia, etnohistoria y antropología: revisión y aportes para el estudio de la religión. *Cuadernos de Teología*, 16, 1–19. <https://doi.org/10.22199/issn.0719-8175-6208>
- Bruun, G. (1964). *La Europa del siglo XIX. 1815-1914*. Fondo de Cultura Económica.
- Cantera, F., & Iglesias, M. (2009). *Sagrada Biblia*. Biblioteca de Autores Cristianos.
- Cappelletti, A. (2007). *Prehistoria del anarquismo*. Libros de la Araucaria.
- Casales, R., & Bastos, L. (2022). *Dios y la filosofía. Una aproximación histórica al problema de la trascendencia*. Tirant.
- Croce, B. (1965). *History of Europe in the nineteenth century*. Unwin University Books.
- D'Angelo, V. (2019). “Ni dieux ni maitre”. Anarquismo y teología política. *Res Publica. Revista de Historia de las Ideas Políticas*, 22(1), 123–140. <https://doi.org/10.5209/rpub.63888>
- Ellul, J. (2011). *Anarchy & Christianity*. Wipf & Stock.
- Hobsbawm, E. (2011). *La era de la revolución. 1789-1848*. Libros de Historia.
- Jiménez García, H. (2021). La crítica al concepto de Dios y sus consecuencias políticas en el pensamiento anarquista post Feuerbachiano y post hegeliano: Mijaíl Bakunin y Max Stirner. *Revista Internacional de Pensamiento Político*, 16, 493–516.
- Kneper, G. (2015). *Bakunin. Biografía global de un libertario cosmopolita*. Universitat Pompeu Fabra.
- Mac Laughlin, J. (2017). *Kropotkin y la tradición intelectual anarquista*. Ariel.
- Maíz, J. (2021). Prólogo: Bakunin contra Dios y contra el Estado. Una lucha incesante. Alianza Editorial.
- Martinic, Z. (2014). Liberalismo e Iglesia en Europa durante el siglo XIX. *Tradición y Saber*, 11(1), 81–99.
- McLaughlin, P. (2002). *Mikhail Bakunin: The philosophical basis of his anarchism*. Algora Publishing.
- McLean, R. (2010). *The Enlightenment*. The Higher Education Academy. University of Warwick.
- Mintz, F. (2006). *Bakunin. Crítica y acción*. Libros de Anarres.
- Outram, D. (2009). *La Ilustración*. Siglo XXI.
- Paniagua, J. (2012). *Breve historia del anarquismo*. Nowtilus.
- Pettazzoni, R. (1974). *L'essere supremo nelle religioni primitive*. Piccola Biblioteca Einaudi.

- Pettazzoni, R. (1996). El ser supremo: estructura fenomenológica y desarrollo histórico. En M. Eliade & J. Kitagawa (Eds.), *Metodología de la historia de las religiones* (pp. 86–94). Paidós Orientalia.
- Piñero, A. (2018). *Aproximación al Jesús histórico*. Trotta.
- Piñero, A. (2020). *El Jesús histórico. Otras aproximaciones*. Trotta.
- Proudhon, P. J. (1975). *¿Qué es la propiedad? Investigaciones acerca del principio del derecho y de la autoridad*. Ediciones Orbis.
- Proudhon, P. J. (2011). *Escritos federalistas*. Akal.
- Stearns, P. (2012). *Una nueva historia para un mundo global*. Crítica.
- Tarleton, G. (1993). *Birth of a Christian anarchist*. Pendragon.
- Van der Leeuw, G. (1931). Die Struktur der Vorstellung des sogenannten höchsten Wesens. *Universitas Nusantara PGRI Kediri*, 29, 79–107.
- Wunn, I. (2012). *Las religiones en la prehistoria*. Akal.
- Xu, M. (2024). The European pattern after the French Revolution: The possibility of historical deduction. *International Journal of Education and Humanities*, 12(3), 45–49. <https://doi.org/10.54097/eayhbm86>
- Zorrilla, R. (1988). Origen y desarrollo del sindicalismo. *Revista Libertas*, 8, 1–13.

Colloquia (colloquia.uhemisferios.edu.ec) is an academic journal of culture and thought of the Universidad de Los Hemisferios. Colloquia is an international publication with a global scope, committed to humanistic education, through the dissemination of articles that are original and relevant for the Humanities. In addition, this journal pursues to advance academic research in Ecuador and the rest of the world. Thanks to its interdisciplinary approach, Colloquia fosters the unity among all types of human knowledge, especially the human sciences.