

Esta publicación se enmarca en las iniciativas académicas de la Escuela de Estudios de Graduados de la Universidad de Heidelberg (Heidelberger Graduiertenschule für Geistes- und Sozialwissenschaften), desarrolladas dentro del Programa de Cooperación Sur-Sur-Norte activo desde el otoño de 2018. La HGGGS tiene como objetivo principal fomentar, estructurar, promover y supervisar programas doctorales interdisciplinarios en las seis facultades que la conforman: Lenguas Modernas, Filosofía, Teología, Derecho, Estudios Culturales Empíricos y del Comportamiento, y Ciencias Económicas y Sociales. El enfoque innovador de la escuela integra metodologías tradicionales con perspectivas transdisciplinarias emergentes, creando puentes entre las humanidades clásicas y los desafíos contemporáneos. Paralelamente, la escuela gestiona iniciativas clave en el ámbito de las humanidades y ciencias sociales, incluyendo proyectos de internacionalización académica, programas de intercambio estudiantil e investigador, y especialmente la gestión de fondos para apoyar la formación doctoral de estudiantes internacionales en Heidelberg. Estos esfuerzos se complementan con una red global de colaboraciones institucionales que enriquecen el diálogo académico entre distintas regiones del mundo. En este contexto, la HGGGS —en coordinación con la oficina de relaciones internacionales— implementa un programa de becas para doctorantes de Bolivia, Cuba y Venezuela, financiado por la Fundación Baden-Württemberg y la propia universidad. La selección de beneficiarios sigue rigurosos criterios de excelencia académica y potencial de impacto social en sus países de origen. El presente volumen recoge los trabajos de seis investigadores del séptimo grupo de becarios que junto a otros autores invitados presentan un análisis interdisciplinario y pedagógico de la Inteligencia Artificial. Cada capítulo refleja tanto la diversidad de enfoques disciplinares como la unidad metodológica que caracteriza a los proyectos de la HGGGS. Esta publicación no solo documenta resultados investigativos, sino que también testimonia el éxito de un modelo de cooperación académica que trasciende fronteras geográficas y culturales.



Reflexión Crítica sobre la IA Astrid Wind, Cástor David Mora y Daniel Quintero (Eds.)

HGGGS

Reflexión Crítica sobre la Inteligencia Artificial

Interpretación interdisciplinaria y pedagógica sobre el escenario tecnológico actual



Reflexiones en torno a los trabajos de investigación desarrollados durante el año 2025 por el programa de cooperación Bolivia, Cuba, Venezuela y Alemania

Astrid Wind, Cástor David Mora y Daniel Quintero (Eds.)



Grupo de Estudios Contemporáneos sobre Digitalidad e Historias Disruptivas (GECDHD)



HGGGS

HEIDELBERGER
GRADUIERTENSCHULE
FÜR GEISTES- UND
SOZIALWISSENSCHAFTEN



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Astrid Wind, Cástor David Mora, Yanett Coromoto Segovia, Daniel Quintero Rodríguez, Nathalie Miret González, Adrián Navia Segovia, Vanessa Márquez Vargas, Alberto Carballo Soca, Debby Avendaño Sánchez y Beatriz Chuquimia Aranda

Reflexión Crítica sobre la Inteligencia Artificial

Interpretación interdisciplinaria y pedagógica sobre el escenario tecnológico actual



Astrid Wind, Cástor David Mora y Daniel Quintero (Editores)



Vista de Heidelberg en Alemania



Vista de Mérida en Venezuela



Vista del Pico Bolívar, Mérida en Venezuela



Vista del Illimani en Bolivia



Vista de La Paz en Bolivia



Vista de La Habana en Cuba



Universidad de La Habana en Cuba



Séptimo grupo en la Universidad Heidelberg



Séptimo grupo en la Universidad Heidelberg



Universidad de Los Andes en Mérida, Venezuela



Frankfurt am Main, Alemania

El presente libro honra a los pensadores del Sur Global, que con análisis críticos fomentan estudios arraigados en ideales comunitarios y una ética solidaria, orientados al beneficio colectivo

Reflexión Crítica sobre la Inteligencia Artificial

Interpretación interdisciplinaria y pedagógica sobre el escenario tecnológico actual

Artículos científicos del séptimo grupo de Cooperación Internacional de la Escuela de Graduados de Ciencias Sociales y Humanidades de la Universidad de Heidelberg (HGGS)

Astrid Wind, Cástor David Mora y Daniel Quintero (Editores)

Edición y diseño: Dr. Cástor David Mora y Prof. Daniel Quintero

Redacción y contenidos: Responsabilidad de cada autor/a

Revisión y lectura final: Dr. Cástor David Mora y Prof. Daniel Quintero

Diagramación: Dr. Cástor David Mora y Prof. Daniel Quintero

Diagramación de tapas: Dr. Cástor David Mora y Prof. Daniel Quintero

© 2025 Astrid Wind, Cástor David Mora, Daniel Quintero y demás autores/as

Primera Edición: 2025

ISBN: 9 783982 219691

Impreso en Heidelberg, Bundesrepublik Deutschland

Agradecimiento (Dankeschön): Gracias a la colaboración de la Baden-Württemberg-STIPENDIUM, a la Oficina de Relaciones Internacionales de la Ruprecht-Karls-Universität Heidelberg, a la Heidelberger Graduiertenschule für Geistes- und Sozialwissenschaften (HGGS) y a todos/as quienes lograron escribir sus aportes para la elaboración del presente libro en el marco de la cooperación Sur-Sur-Norte.

Aviso legal: La Ruprecht-Karls-Universität Heidelberg (UH), la Heidelberger Graduiertenschule für Geistes- und Sozialwissenschaften (HGGS) y los editores del presente libro no se hacen responsables por ninguna información incompleta, incorrecta, plagiada o elaborada por „máquinas inteligentes“, así como tampoco por las inconsistencias, faltas u omisiones en la citación de las respectivas referencias bibliográficas, hemerográficas y electrónicas. Los contenidos de esta publicación expresan el punto de vista personal de los autores, quienes son los únicos responsables de sus escritos y son divulgados con el propósito de generar una reflexión crítica sobre la *Inteligencia Artificial*. De ningún modo debe entenderse que los mismos representan la posición de la UH, la HGGS o los editores.

Portada: Ekeko boliviano. © 2010 Mamani Mamani.

Imágenes de las páginas preliminares: aportadas por los autores y el equipo editorial. Las fotografías de Cuba y de la ULA fueron descargadas de Wikipedia. **Créditos:** José Alfaro (2015), Andreas Wieland (2016) y Nino Verde (2013).

Índice

Prólogo.	
Los editores	10
Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica.	
Astrid Wind y Cástor David Mora	15
El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo.	
Yanett Coromoto Segovia	108
Reflexión crítica sobre el uso hegemónico de la IA para la posverdad.	
Daniel Quintero Rodríguez	132
Desmitificando la atribución de responsabilidad penal por actuaciones de la IA: perspectivas desde la robótica.	
Nathalie Miret González	155
IA y hegemonía en disputa: China, los BRICS y el Sur global en la reconfiguración del poder tecnológico.	
Adrián Navia Segovia.....	194
Valoración crítica de la interacción Humano-IA en la construcción de sentido: Ejercicio en el marco de la Semiótica de las Pasiones.	
Vanessa Márquez Vargas	216
Producción científica sobre alfabetización en inteligencia artificial: estudio bibliométrico en Web of Science.	
Alberto Carballo Soca.....	241
Narrativas Distópicas y Utópicas: Uso de la IA en la Arquitectura.	
Debby Avendaño Sánchez	272
Inteligencia Artificial: Análisis de su impacto en la brecha digital.	
Beatriz Chuquimia Aranda	285
Equipo editorial, las autoras y los autores.....	315

Prólogo

Esta publicación surge como parte de las iniciativas académicas impulsadas por la Escuela de Estudios de Graduados de la Universidad de Heidelberg (Heidelberger Graduiertenschule für Geistes— und Sozialwissenschaften HGGS), en el marco de su Programa de Cooperación Sur-Sur-Norte establecido desde 2018. Como eje central de formación doctoral, la HGGS articula esfuerzos interdisciplinarios entre seis facultades: Lenguas Modernas, Filosofía, Teología, Derecho, Estudios Culturales Empíricos y del Comportamiento, y Ciencias Económicas y Sociales, combinando metodologías tradicionales con enfoques innovadores que conectan las humanidades con los desafíos globales actuales. La escuela ha desarrollado una sólida plataforma de internacionalización académica que incluye intercambios estudiantiles, cooperación investigativa y un programa especial de becas para doctorantes de Bolivia, Cuba y Venezuela, financiado conjuntamente por la Fundación Baden-Württemberg y la Universidad de Heidelberg. Este programa —que prioriza la excelencia académica y el potencial de impacto social— ha permitido la formación de una nueva generación de investigadores cuyos trabajos, junto con contribuciones de autores invitados, conforman el núcleo del presente volumen. Los capítulos aquí reunidos ejemplifican el modelo colaborativo de la HGGS, donde la diversidad disciplinar (desde el derecho hasta los estudios culturales) se armoniza con un riguroso enfoque metodológico compartido. Más allá de sus aportes específicos sobre inteligencia artificial, esta obra representa los frutos de un esquema de cooperación académica que supera barreras geográficas, demostrando cómo el diálogo Sur-Sur-Norte puede generar conocimiento transformador ante los desafíos tecnológicos contemporáneos. La publicación no solo difunde resultados investigativos, sino que también consolida una red de pensamiento crítico transnacional comprometida con la justicia cognitiva global.

El comprender la IA como un desafío civilizatorio amerita miradas desde el Sur Global y Europa para interpretar el escenario tecnológico actual. Mientras que en Alemania por intermedio del Deutscher Ethikrat se impulsa una discusión ética y regulatoria, las naciones del Sur Global enfrentan una compleja paradoja: adoptar estas tecnologías para no quedar rezagadas, pero sin caer en nuevas formas de dependencia neocolonial. Esta obra colectiva reúne nueve perspectivas fundamentales que van desde la pedagogía crítica hasta el análisis geopolítico, desmontando el discurso hegemónico que impera en el planeta. Los autores aquí reunidos —Astrid Wind, Cástor David Mora, Yanett Coromoto Segovia, Daniel Quintero Rodríguez, Nathalie Miret González, Adrián Navia Segovia, Vanessa Márquez Vargas, Alberto Carballo Soca, Debby Avendaño Sánchez y Beatriz Chuquimia Aranda— coinciden en una reflexión crucial: la IA no es meramente un conjunto de herramientas técnicas, sino un campo de batalla donde se juega el futuro de la autonomía educativa,

la justicia social, la soberanía tecnológica y sobre todo la esencia de lo humano.

La obra inicia con un aporte de Astrid Wind y Cástor David Mora: „Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica“, ofreciéndonos una reflexión fundamental sobre los impactos de la IA en los procesos educativos, desmontando meticulosamente el discurso tecnocrático dominante. Su análisis histórico, que va desde los fundamentos lógicos de Alan Turing hasta las redes neuronales contemporáneas, revela cómo la llamada „inteligencia“ artificial carece de los elementos esenciales que caracterizan al pensamiento humano: conciencia, contextualización y capacidad crítica. La propuesta del hexágono pedagógico representa un avance teórico crucial al expandir el modelo tradicional de enseñanza, incorporando dimensiones éticas, interculturales y críticas que los sistemas algorítmicos no pueden abordar. Ambos autores alertan sobre riesgos concretos como la creciente descualificación docente y la mercantilización del conocimiento, proponiendo en cambio una apropiación crítica de estas tecnologías al servicio de proyectos educativos emancipadores. Su distinción entre aprendizaje humano (contextual, significativo) y aprendizaje automático (estadístico, acrítico) constituye un llamado urgente a reivindicar la educación como proceso esencialmente humano.

Más adelante, Yanett Coromoto Segovia con „El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo“, establece una profunda conexión entre el auge de la Inteligencia Artificial y la transformación, casi extinción, de la tradición oral de la etnia wayuu en entornos urbanos, utilizando como marco teórico la lúcida propuesta de Walter Benjamin sobre la desaparición del narrador en la modernidad. Los wayuu, un pueblo indígena de la península de La Guajira, en la frontera entre Colombia y Venezuela, forjaron durante siglos su complejo universo simbólico a través de una sofisticada ritualización de la palabra, categorizada como „fría“, „caliente“ o „que se lleva“, y encarnada en la figura fundamental del „Püpchipü“, un mediador y narrador ancestral. En una paradoja típicamente moderna, justo cuando el narrador tradicional desaparece, surge una fascinación melancólica por „lo perdido“, por aquella forma auténtica que se evapora. Benjamin, en su célebre texto de 1936, ya había avizorado cómo la modernidad tecnológica e industrial erosionaba la „Erfahrung“, esa experiencia colectiva y acumulativa que se transmitía mediante los relatos, sustituyéndola por el flujo de información descontextualizada y efímera. Casi un siglo después, la inteligencia artificial emerge como el epítome de este proceso: un „narrador invisible“ que genera discursos fluidos, masivos y carentes de toda corporalidad, e incluso aspira a asumir funciones de cuidado simbólico sobre la vida y la conciencia. Lejos de ser una ruptura, la IA representa la radicalización máxima de esa „pobreza de experiencia“ benjaminiana, pues su

Equipo editorial

narrativa es un producto algorítmico, exento de historia natural, de memoria encarnada y de la autoridad que da la vivencia real.

Seguidamente, se presenta el escrito de Daniel Quintero Rodríguez „Reflexión crítica sobre el uso hegemónico de la IA para la posverdad“ que desarrolla una crítica profunda al papel de la IA como herramienta de dominación en el capitalismo cognitivo contemporáneo. Inspirado en el marco gramsciano, el autor analiza cómo los „Centros de Gestión Artificial de la Posverdad“ (CGAP) utilizan algoritmos para construir alternativas a la realidad que favorecen un „cautiverio intelectual“ donde las plataformas digitales aíslan a los sujetos en burbujas informativas, impidiendo la formación de pensamiento crítico colectivo. Quintero revela la cadena de dominación capitalismo cognitivo + tecno-liberalismo, mostrando cómo la IA neutraliza la acción política contrahegemónica. Frente a esto, propone el Contrabloque Histórico Insurgente, sostenido por una „Filosofía de la Praxis“ que recupere la historicidad del sujeto y se enfrente a temas complejos como la explotación cognitiva. Su análisis es particularmente relevante para el Sur Global, donde la batalla por la soberanía comunicacional adquiere dimensiones estratégicas.

Continúa la disertación de Nathalie Miret González „Desmitificando la atribución de responsabilidad penal por actuaciones de la IA: perspectivas desde la robótica“ donde la autora aborda con rigor jurídico uno de los problemas más complejos de la IA contemporánea: la creciente autonomía de los sistemas inteligentes y sus implicaciones legales. A través del análisis de modelos como MITL y HOOTL, expone los desafíos que surgen cuando sistemas de IA toman decisiones sin supervisión humana. La autora introduce el debate sobre „personalidad electrónica“, demostrando cómo esta ficción podría eximir de responsabilidad a algunos actores tecnológicos. Partiendo de una metodología cualitativa y un enfoque interdisciplinario se revelan las limitaciones de los marcos legales diseñados para acciones humanas. Sus propuestas analíticas son pertinentes en el contexto Latinoamericano y Caribeño, cuyos sistemas jurídicos enfrentan el desafío de regular tecnologías desarrolladas en contextos ajenos a sus realidades nacionales.

Luego se encuentra la propuesta de Adrián Navia Segovia: „IA y hegemonía en disputa: China, los BRICS y el Sur global en la reconfiguración del poder tecnológico“ que examina a la IA como un espacio central de competencia por la dominación tecnológica a nivel mundial, destacando el crecimiento de China y la relevancia de los BRICS. El autor utiliza un marco teórico que integra las perspectivas de Arrighi, Wallerstein y Harvey, para valorar cómo la IA manifiesta y fortalece las relaciones de poder internacional. Además, el artículo evalúa los procesos geopolíticos vinculados a la influencia tecnológica china y la colaboración entre países del Sur (especialmente en Latinoamérica), considerando posibles dinámicas de dependencia y colonialismo digital. Para Navia, la IA se presenta como un campo de batalla política, económica y

cultural donde se reconfigura el equilibrio global de poder, siendo necesario que el Sur global genere su propio conocimiento tecnológico en lugar de limitarse a consumirlo, debiendo promover una gobernanza digital diversa con una IA ética y autónoma.

En el meridiano de la obra se encuentra la reflexión de Vanessa Márquez Vargas „Valoración crítica de la interacción Humano-IA en la construcción de sentido: Ejercicio en el marco de la Semiótica de las Pasiones“, quien desarrolla una crítica fundamental desde la semiótica a la pretensión de que los modelos de lenguaje „comprenden“ significados. Mediante análisis comparativos de respuestas de diversos sistemas de IA, demuestra que estos solo simulan comprensión mediante patrones estadísticos, careciendo de verdadera interpretación significativa. Partiendo del concepto de „borde semiótico“ describe el límite infranqueable entre simulación algorítmica y comprensión humana arraigada en la experiencia encarnada. La autora insiste en la necesidad de preservar la dimensión pasional y hermenéutica del sujeto humano frente a la creciente mediación tecnológica de la comunicación.

En cuanto al aporte de Alberto Carballo Soca „Producción científica sobre alfabetización en inteligencia artificial: estudio bibliométrico en Web of Science“ analiza 268 documentos sobre alfabetización en IA, revelando sus hallazgos un crecimiento exponencial del interés académico en el tema, tras un período de desarrollo inicial en 2019, alcanzó su punto máximo entre 2023 y 2024. Explica Cabello que la producción se concentra en un reducido grupo de revistas especializadas, siendo Education and Information Technologies la que más publicaciones realiza. Se describe en el texto que un núcleo muy activo de autores tiene una alta tasa de citación, mientras que la mayoría de los investigadores participa de forma ocasional, un rasgo típico de una disciplina joven. Geográficamente, instituciones de Asia Oriental, Norteamérica y Europa son las más productivas, con China y EE.UU. a la cabeza tanto en publicaciones como en cooperación internacional. Las palabras clave más usadas, inteligencia artificial y alfabetización en inteligencia artificial, enfatizan la urgencia de establecer fundamentos educativos bien definidos para enseñar estas competencias en el Sur Global.

En relación al estudio de Debby Avendaño Sánchez „Narrativas Distópicas y Utópicas: Uso de la IA en la Arquitectura“ se examinan las visiones opuestas respecto a la implementación de la inteligencia artificial en arquitectura, considerándola tanto una herramienta beneficiosa para el diseño como un riesgo por sus consecuencias éticas, sociales y ontológicas. La metodología emplea un análisis crítico-discursivo, complementado con un estudio histórico-cultural de las narrativas tecnológicas en este campo. Mediante una revisión documental exhaustiva, se investigan casos relevantes, nociones filosóficas y representaciones mediáticas que construyen percepciones utópicas y distópicas. La investigación se organiza en dos perspectivas narrativas

Equipo editorial

principales: aquella que promueve la adopción tecnológica y aquella que alerta sobre sus efectos deshumanizantes. Los hallazgos de la autora exponen conflictos entre la eficiencia de los algoritmos y la creatividad humana, así como peligros de despersonalización laboral, estéticas artificiales, deterioro profesional y vaciamiento disciplinario. Pese a ello, también se identifican contribuciones positivas en visualización, diseño asistido e innovación conceptual. Cabe resaltar que las posturas críticas, más allá de señalar problemas, históricamente han impulsado reflexiones y rejuvenecimiento teórico en la arquitectura.

Para el cierre, Beatriz Chuquimia Aranda expone su trabajo „Inteligencia Artificial: Análisis de su impacto en la brecha digital“ que evidencia cómo la IA está exacerbando las desigualdades estructurales en América Latina. Plantea la autora que esta brecha ha evolucionado, dejando de ser solo una cuestión de acceso para convertirse en un problema multifacético que incluye diferencias en habilidades, condiciones de uso, territorio, economía y género. El escrito analiza específicamente el impacto de la IA en esta brecha, utilizando datos socio-demográficos sobre acceso a banda ancha y costos en Bolivia y América Latina. El estudio concluye que el modelo de negocio basado en suscripciones de pago adoptado por las grandes corporaciones tecnológicas agrava las desigualdades económicas, ya que privilegia a los países y sectores con mayor poder adquisitivo, dejando en una clara desventaja a las regiones y comunidades con menos recursos, ampliando así la brecha digital existente debido a los altos costos asociados.

Esta obra colectiva trasciende la mera problematización para convertirse en un manifiesto urgente: frente al avance acrítico de la inteligencia artificial, los diez autores trazan caminos concretos para construir una relación soberana y humanista con estas tecnologías. Los capítulos demuestran que otra IA es posible —descentralizada, al servicio de los pueblos y sujeta a control democrático— pero su construcción requerirá movilizar el pensamiento crítico, fortalecer la educación pública, desarrollar marcos legales audaces y tejer alianzas sur-sur que eviten bloques hegemónicos digitales. Más que una tecnología, la IA es espejo que refleja nuestras contradicciones civilizatorias: en el que debemos mirarnos con valentía para decidir colectivamente qué mundo queremos habitar en el futuro.

Equipo editorial

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica.

Artificial Intelligence, Learning, and Teaching: A Critical Perspective.

Astrid Wind
astrid.wind@hggs.uni-heidelberg.de

Cástor David Mora
david.mora@uni-heidelberg.de

Resumen

Este documento ofrece un análisis crítico, interdisciplinario y pedagógico sobre la Inteligencia Artificial (IA), abordando su evolución histórica, fundamentos lógicos y matemáticos, así como sus impactos actuales en el ámbito educativo, social, científico y cultural. Se inicia con una reflexión sobre la relación entre IA, aprendizaje y enseñanza, destacando sus potencialidades, pero también sus riesgos. A nivel histórico, se revisa el papel de Alan Turing y la influencia de tres grandes matemáticos en la formulación de la lógica computacional. Se analizan el Test de Turing y la crítica de la Habitación China como puntos clave en la discusión sobre la inteligencia de las máquinas. El texto explora el desarrollo de las redes neuronales artificiales, su estructura, analogía con la neurona biológica y aplicaciones destacadas como en el campo de la medicina. Asimismo, se presentan diversas posturas a favor y en contra de la expansión masiva de la IA en todos los ámbitos de la vida, junto con un análisis reflexivo sobre sus inconvenientes. Especial atención se dedica al impacto de la IA en la educación, proponiendo una transición del triángulo pedagógico tradicional al hexágono pedagógico de Wind y Mora, que incorpora nuevas dimensiones éticas, tecnológicas, interculturales e interdisciplinarias, críticas y transformadoras. Se discuten los riesgos de descualificación profesional, la sobrecarga de trabajo docente y la pérdida de cualificaciones académicas frente a la automatización. Finalmente, se plantea una distinción esencial entre el aprendizaje humano y el aprendizaje automático, reafirmando el valor irreductible de la inteligencia humana en contextos mediados por tecnología.

Palabras claves: análisis, crítico, interdisciplinario, pedagógico, inteligencia artificial.

Abstract

This document offers a critical, interdisciplinary, and pedagogical analysis of Artificial Intelligence (AI), addressing its historical evolution, logical and mathematical foundations, as well as its current impacts in the educational, social, scientific, and cultural spheres. It begins with a reflection on the relationship between AI, learning, and teaching, highlighting its potentialities, but also its risks. Historically, it reviews the role of Alan Turing and the influence of three great mathematicians in the formulation of computational logic. The Turing Test and the Chinese Room critique are analysed as key points in the discussion on machine intelligence. The text explores the development of artificial neural networks, their structure, their analogy with biological neurons and notable applications such as in the field of medicine. It also presents various positions for and against the massive expansion of AI in all areas of life, together with a reflective analysis of its drawbacks. Special attention is paid to the impact of AI on education, proposing a transition from the traditional pedagogical triangle to the pedagogical hexagon of Wind and Mora, which incorporates new ethical, technological, intercultural and interdisciplinary, critical and transformative dimensions. The risks of professional disqualification, teacher overload and the loss of academic qualifications in the face of automation are discussed. Finally, an essential distinction is made between human learning and machine learning, reaffirming the irreducible value of human intelligence in technology-mediated contexts.

Keywords: analysis, critical, interdisciplinary, pedagogical, artificial intelligence.

1. Introducción

La inteligencia artificial (IA) ha dejado de ser una promesa futura para consolidarse como una realidad omnipresente que redefine las estructuras del conocimiento, las formas del trabajo, los vínculos sociales, los modos de enseñanza y aprendizaje. Su impacto en la vida cotidiana, la ciencia, la industria, la salud y la educación es cada vez más profundo y transversal. Sin embargo, a pesar de su creciente relevancia, la IA no puede ser comprendida únicamente desde una perspectiva técnica o instrumental. Requiere, por el contrario, un abordaje crítico, multidimensional y transdisciplinario que permita identificar no sólo sus beneficios y oportunidades, sino también sus limitaciones, contradicciones y riesgos, los cuales realmente son muchos. Este documento, estructurado en un conjunto amplio de secciones, se propone

ofrecer una reflexión sistemática y rigurosa sobre la IA, desde sus fundamentos históricos, matemáticos y lógicos hasta sus implicaciones filosóficas, pedagógicas y sociopolíticas.

En primer lugar, se propone una mirada crítica al vínculo entre inteligencia artificial, aprendizaje y enseñanza. Se advierte que, lejos de tratarse de una simple herramienta auxiliar, la IA está modificando los marcos epistémicos y didácticos que sustentan los procesos educativos. La promesa de personalización del aprendizaje, eficiencia evaluativa y acceso a contenidos globales convive con la amenaza de una creciente automatización de funciones cognitivas que han sido, históricamente, propias del docente y del estudiante. En este contexto, se hace necesario contextualizar el fenómeno de la IA desde una perspectiva histórica, crítica y transformadora.

La revisión de su breve pero intensa trayectoria permite comprender los orígenes del pensamiento computacional moderno, el papel decisivo de Alan Turing y la influencia de grandes matemáticos como Gödel, Hilbert y Boole en la concepción de una lógica formal capaz de sustentar procesos de cálculo simbólico.

La obra de Turing es abordada no sólo por su valor histórico, sino por su centralidad teórica. El análisis del Test de Turing permite abrir el debate sobre la posibilidad de que las máquinas puedan pensar, aprender y responder como un ser humano. Esta cuestión, lejos de ser puramente técnica, es fundamentalmente filosófica y epistemológica. Por ello, se incluye el estudio de la Habitación China de John Searle, una de las críticas más influyentes al test de Turing y a la IA en general, que pone en tela de juicio la idea de comprensión semántica por parte de los sistemas computacionales. La discusión se enriquece al contrastar la interpretación mecanicista del procesamiento de símbolos con la concepción intencional y contextual de la mente humana.

Avanzando en el recorrido, se examina el surgimiento de las redes neuronales artificiales, explicando su estructura y funcionamiento desde una analogía biológica. La descripción del modelo de la neurona biológica y su traducción al modelo de la neurona artificial permite entender cómo la IA moderna simula ciertos aspectos del aprendizaje humano mediante capas de procesamiento interconectadas.

Se muestra cómo estas redes han sido fundamentales para el desarrollo de aplicaciones concretas, como los sistemas de diagnóstico médico asistido por IA, que representan uno de los casos más significativos de su utilidad en la práctica profesional. Sin embargo, esta funcionalidad técnica no puede ocultar la necesidad de una regulación crítica y contextualizada de su uso, pero también el análisis crítico social, político, técnico y formativo de aspectos

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

fundamentales que están directamente relacionados con el uso masivo de la IA.

El texto también expone los argumentos a favor de la utilización masiva de la IA en múltiples sectores: desde la educación hasta la industria, pasando por la cultura, la comunicación y la ciencia. A su vez, presenta una contrapartida reflexiva: los posicionamientos en contra, centrados en los efectos negativos de la automatización desmedida, como la vigilancia, el desplazamiento laboral, la despersonalización de los procesos formativos y la dependencia tecnológica. A partir de estos enfoques, se desarrollan los principales inconvenientes de la IA, no como elementos excluyentes de su desarrollo, sino como advertencias que requieren atención ética, política y psicopedagógica.

Especial atención merece el bloque dedicado a la IA en el contexto pedagógico. Se cuestiona la viabilidad del clásico triángulo pedagógico en un entorno dominado por algoritmos, y se propone en su lugar el hexágono pedagógico de Astrid Wind y Cástor David Mora, que incorpora nuevas dimensiones fundamentales para afrontar los desafíos actuales: la dimensión social, cultural, ecológica, humana, económica y tecnológica. Esta reformulación pedagógica se vincula estrechamente con el análisis de los principales desafíos de la integración de la IAG en los procesos didácticos, que incluyen la redefinición de los roles del docente y del estudiante, el rediseño curricular y la necesidad de políticas públicas inclusivas e innovadoras.

Las últimas secciones del documento desarrollan, con especial énfasis, el problema de la descualificación como una de las consecuencias más graves y menos visibilizadas de la IA. Se analiza la manera en que la IAG puede generar sobrecarga de trabajo, pérdida de saberes prácticos, desplazamiento de funciones cognitivas esenciales y debilitamiento del pensamiento crítico. Se identifican los elementos característicos de la pérdida potencial de cualificaciones humanas, especialmente en el ámbito universitario, donde la función docente puede verse sustituida por sistemas de tutoría automatizados, evaluación algorítmica y producción de contenido generativo. En este contexto, la pregunta por la diferencia entre inteligencia humana e inteligencia artificial se vuelve urgente, no sólo desde una perspectiva técnica, sino desde una ética de la educación, de los saberes y los conocimientos.

El documento culmina con una reflexión sobre el aprendizaje humano frente al aprendizaje automático, destacando las capacidades de significación, empatía, juicio y creatividad que aún distinguen al ser humano frente a los sistemas artificiales. En suma, este trabajo no busca ofrecer respuestas definitivas, sino abrir un espacio de discusión rigurosa y crítica que permita repensar el lugar de la inteligencia artificial en nuestras vidas, en nuestras aulas y en nuestras instituciones. La IA no es una amenaza inevitable ni una

solución mágica, sino un campo en disputa que debe ser analizado, regulado y reconducido a partir de criterios democráticos, críticos, políticos, educativos, humanistas y transformadores.

2. Preliminares

El presente trabajo tiene por finalidad ocuparse del tema de la llamada Inteligencia Artificial (IA), el cual comprende un sinnúmero de aspectos que sería muy difícil abarcarlo en un libro y mucho menos en un artículo de opinión crítica como el que tienes en tus manos. La IA es considerada actualmente como una „tecnología revolucionaria“ que determinará, en un futuro muy cercano, prácticamente la totalidad de la vida de los seres humanos, independientemente de la región geográfica donde se encuentre (Brynjolfsson y McAfee, 2014; Häggström, 2016; Müller y Bostrom, 2016; Spiteri, 2023; OEI, 2023; Gawdat, 2024; Lozano y Garibay, 2024; Santana, 2024; entre muchos/as otros/as). Hoy la llamada IA es considerada como la panacea tecnológica de la actualidad y del futuro¹. Pocas veces una temática ha tenido tanto interés científico, tecnológico, político, social, económico, cultural, lingüístico y popular, con una cobertura mediática inmensa durante un período de tiempo tan corto (aproximadamente tres años) como ha ocurrido con la denominada Inteligencia Artificial. La abundancia de literatura sobre este aspecto, sólo sería comparable con la abundancia de escritos surgidos a raíz de la pandemia del coronavirus en marzo del 2020 (Wind y Mora, 2023a). Es muy probable que este mar de opiniones generales, trabajos científicos, escritos técnicos académicos y reportes de investigaciones que aparecen con tanta frecuencia, tenga que ver con el apoyo de los „softwares inteligentes“ que estarían ayudando a las personas, de una u otra manera, a la producción de tales escritos.

En otros trabajos hemos descrito con cierto detalle el desarrollo histórico de la computación e informática en general (Mora y Wind, 2023a; Heßler, 2025), especialmente en el marco del desarrollo de la tecnología computacional, gracias a los aportes esencialmente de las matemáticas. Estos avances de la computación han traído como resultado concreto la conformación de lo que se conoce actualmente como Inteligencia Artificial, lo cual no es más que la creación de „máquinas“, que tiene como objetivo la realización de diversa tareas propias de las potencialidades humanas, tales como comprender y producir textos de toda naturaleza, realizar imágenes, producir simulaciones de toda naturaleza, tomar decisiones, aprender autónoma e independientemente, razonar de acuerdo con situaciones sencillas o complejas, centrarse en mecanismos de percepción y discernimiento, etc.

¹ **Nota de los Autores:** Para ver con mayor detalle el desarrollo histórico de la tecnología y las revoluciones industriales, recomendamos consultar los trabajos de Cástor David Mora y Astrid Wind.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

Por supuesto que la idea básica de la IA no es reciente, siempre ha estado asociado al interés y necesidad tecnológica de los seres humanos en cuanto a la búsqueda de herramientas tecnológicas que le permitan resolver situaciones complejas de su mundo de vida de manera eficiente, rápida y precisa. A pesar de los intentos antiguos reales, creativos o imaginarios realizados por los seres humanos con la finalidad de sustituir sus acciones físicas y cognitivas, es sólo durante la segunda mitad del siglo veinte que tiene lugar el inicio de la ampliamente propagada „inteligencia artificial“ (Metzinger, 2015; Stone, Peter; et al., 2016; Mora y Wind, 2023a; Heßler, 2025). Para muchos/as investigadores/as, especialmente en el campo de las matemáticas, el objetivo e interés estuvo siempre centrado en la conformación de máquinas calculadoras automáticas, pensantes e inteligentes que pudiesen no tanto sustituir a los seres humanos, sino más bien ayudarlos a resolver problemas aburridos, reiterativos, complejos, tal como lo hacen las personas con su pensamiento y acción. A pesar de que este avance fue, supuestamente lento para muchas personas, los/as matemáticos e informáticos/as nunca perdieron el horizonte y el objetivo central sobre la conformación de las „máquinas inteligentes“ (Müller, 2023; Robisco, 2024;). Tal vez, la supuesta lentitud de este progreso constante tuvo que ver con dos razones fundamentales, por un lado, los propios adelantos matemáticos propiamente dichos, como por ejemplo la optimización algorítmica y, en segundo lugar, tal vez el más importante, el progreso técnico o para algunos/as las limitaciones materiales inherentes a la implementación de los modelos y algoritmos matemáticos a dichas máquinas.

Estas supuestas barreras fueron rotas definitivamente durante el transcurso del presente siglo. Desde hace unas dos décadas aproximadamente se superó definitivamente las trabas entre las teorías matemáticas que hacían posible el desarrollo de la inteligencia artificial y, por supuesto, los inmensos y veloces servidores, así como el acelerado avance de las redes mundiales de información a través del internet que posibilitaron una gran acumulación de datos e informaciones durante los últimos años del siglo veinte, puesto que no puede haber „inteligencia artificial“, si no existen saberes y conocimientos, datos, evidencias e informaciones, obviamente producidos por los mismos seres humanos durante su propia existencia. Todo ello permitió el desarrollo rápido del conocido aprendizaje profundo, la visión computarizada, el pensamiento mecánico, la codificación y decodificación de las lenguas naturales, los programas de traducción, los cuales fueron desarrollados hace más de tres décadas, entre otras competencias mecánicas, algorítmicas y computacionales.

Es muy importante destacar que entre los años 1948 y 1970, aproximadamente, tuvo lugar un importante movimiento informático mundial, el cual se concentró en el desarrollo de diversos programas que comprendían una gama amplia de propuestas, pasando desde los largamente conocidos

como sistemas de expertos que podrían aprender a jugar juegos de mesa sencillos, pero también otros más complejos como el ajedrez. Durante ese período de tiempo también fueron creados algunos lenguajes de programación como el famoso programa de inteligencia artificial conocido como LISP, el cual consistió en una familia de lenguajes de programación desarrollado en 1958 por el Instituto de Tecnología de Massachusetts (ITM) (Winston y Horn, 1987; Queinnec, 2003). Después del lenguaje Fortran, LIPS es el lenguaje de programación más antiguo que aún es usado en la actualidad. Igualmente, buena parte de la población mundial sabe que la informática está directamente asociada a las matemáticas y, muy especialmente, a la elaboración de algoritmos matemáticos, los cuales obviamente tienen que ver con la conceptualización de conceptos siendo uno de los más importantes el creado por Frank Rosenblatt en 1958 y ampliamente conocido como el Perceptrón; este concepto matemático tiene por objetivo mostrar cómo se percibe la información a partir del mundo físico, de qué manera se puede recordar los datos e informaciones y, muy especialmente, cómo influye dicha información almacenada en la memoria de un sujeto o de una máquina en el reconocimiento y comportamiento de tales sujetos y máquinas.

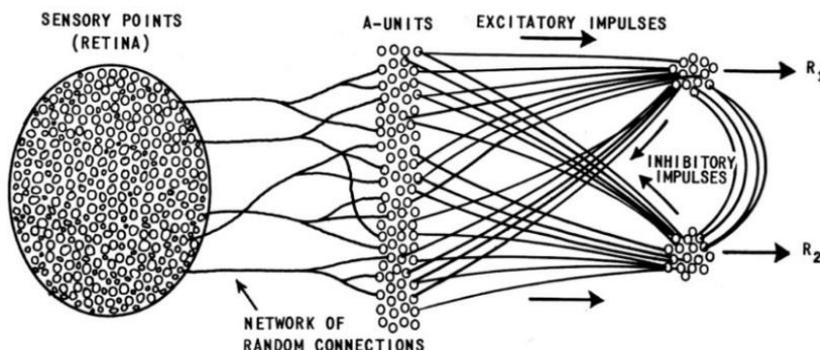
De esta manera el perceptrón consiste en un concepto teórico que permite simular a un sistema nervioso hipotético. Lo importante es que esta teoría, interdisciplinaria por excelencia, admite establecer una relación o un puente entre la biofísica y la psicología, lo cual nos permite elaborar propositivamente (predicción) curvas de aprendizaje a partir de ciertas variables neurológicas y viceversa. Para ello, la teoría del Perceptrón se basa en la manipulación de grandes datos estadísticos cuantitativos que son indispensables para poder comprender la organización de los sistemas cognitivos, especialmente de los seres humanos (Rosenblatt, 1958, 390). Esta unidad fundamental del cálculo se basa en un algoritmo que podía recibir un conjunto de datos de entrada (por ejemplo, mediante la retina), seleccionar un subconjunto de los mismos con base en criterios suministrados por los/as programadores/as. Como se puede apreciar intervienen aquí de manera directa, por lo menos, varias disciplinas científicas; es decir, las matemáticas, la física, la biología, la fisiología y la psicología.

El modelo del perceptrón propuesto por Frank Rosenblatt (1958, 390) representa uno de los primeros intentos exitosos de emular computacionalmente el funcionamiento de las neuronas biológicas. Inspirado en la neurociencia, Rosenblatt diseñó una arquitectura capaz de aprender a clasificar patrones a partir de estímulos sensoriales. En la imagen que aparece en las figuras 1a y 1b, se puede observar cómo los puntos sensoriales (retina) captan la información del entorno, actuando como una capa de entrada que transmite señales a través de una red de conexiones aleatorias hacia una segunda capa de procesamiento denominada unidades A (A-units). Estas A-units están conectadas a su vez con las unidades de salida (R_1 y R_2), y la

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

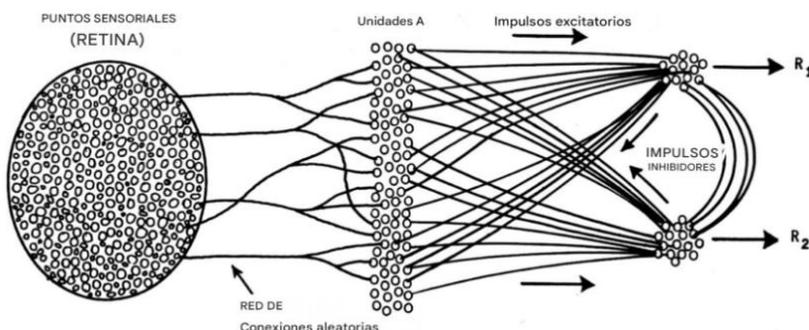
transmisión entre capas ocurre mediante impulsos excitatorios o inhibitorios, que modelan el proceso biológico de activación o supresión neuronal.

Figura 1a: Esquema del Perceptrón de Rosenblatt (1958)



Fuente: Rosenblat (1958, 390).

Figura 1b: Esquema del Perceptrón de Rosenblatt (1958)



Fuente: Rosenblat (1958, 390).

Cada conexión tiene un peso que determina su influencia sobre la salida final. Durante el entrenamiento, el perceptrón ajusta estos pesos con base en la diferencia entre la salida esperada y la salida real, lo que le permite aprender a reconocer patrones como letras, formas geométricas o estímulos visuales simples, de manera autónoma. Este mecanismo fue el precursor del algoritmo de aprendizaje supervisado mediante reglas de refuerzo. El trabajo de Rosenblatt fue revolucionario porque introdujo el concepto de aprendizaje automático basado en experiencias previas, anticipando los fundamentos de las redes neuronales artificiales modernas. Aunque el perceptrón original solo podía resolver problemas linealmente separables —una limitación señalada por Minsky y Papert en 1969—, su diseño marcó el inicio de la Inteligencia Artificial conexionista, que más tarde evolucionaría hacia arquitecturas más

profundas y sofisticadas como el perceptrón multicapa y las redes convolucionales. En suma, el modelo de Rosenblatt representa el eslabón fundacional entre la biología neural y la computación inteligente.

Breve historia de la Inteligencia Artificial (AI)

En la actualidad corren por todos los vientos las palabras Inteligencia Artificial (IA) como si fueran palabras mágicas, salvadoras de la supuesta incapacidad de los seres humanos para seguir creando y produciendo ideas, datos, conceptos, informaciones, cuentos, historias, poesía, tecnología, etc., incluyendo obviamente a la llamada IA. Si bien, muchas personas siguen considerando que la misma tiene que ver con motores o máquinas de búsqueda en internet, asistentes virtuales como Alexa, Siri u otros, aplicaciones que resuelven problemas digitalmente, programas simuladores, traductores, entre otras consideraciones. Sin embargo, pocos/as usuarios/as de esta tecnología se han preguntado por su origen, en qué consiste, cuál es su importancia para los seres humanos, cómo serán nuestras sociedades en un futuro cercano con su implementación masiva o, algo más grave, ¿será posible replicar y sustituir la inteligencia humana mediante la tecnología? Estas interrogantes también fueron formuladas, entre muchos/a otras/as, por Mora y Wind hace dos años (Wind y Mora, 2023a).

Con respecto a los primeros pasos de la IA, nos remitimos aquí a señalar que es precisamente a mediados de la década de los cincuenta del siglo pasado, concretamente en 1955, cuando apareció por primera vez este término en el idioma inglés. Su surgimiento estuvo asociado, como casi siempre ocurre con estas tendencias filantrópicas, a un evento organizado en aquel año por la Fundación Rockefeller, el cual tuvo por finalidad discutir el desarrollo y futuro de los autómatas y modelos cerebrales, basados estos últimos en complejos modelos matemáticos. El documento conclusivo de dicho encuentro contenía terminologías como lenguajes de programación, redes neuronales, abstracción y creatividad artificial, máquinas pensantes, rebotes autónomos, etc.

Este primer trabajo dio como resultado que exactamente un año después el joven matemático, John McCarthy, de la universidad de Dartmouth, conformará un equipo de técnicos, matemáticos, informáticos, tales como Claude Shannon, Marvin Minsky y Ray Solomonoff, el cual tuvo como tarea principal discutir de manera intra e interdisciplinariamente los temas surgidos en el encuentro organizado en 1955 por la Fundación Rockefeller. Durante varios días se reunieron, de manera intermitente, aproximadamente treinta personas provenientes de diversas disciplinas científicas; de allí, su connotación e importancia interdisciplinaria. Este importante taller de trabajo práctico y discusión teórica tuvo lugar en la Universidad de Dartmouth, asumieron las siguientes tres grandes tesis: a) La inteligencia no es exclusiva del cerebro humano; b) La inteligencia puede ser descrita de manera formal;

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

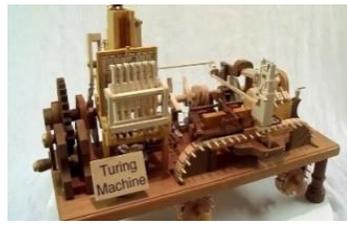
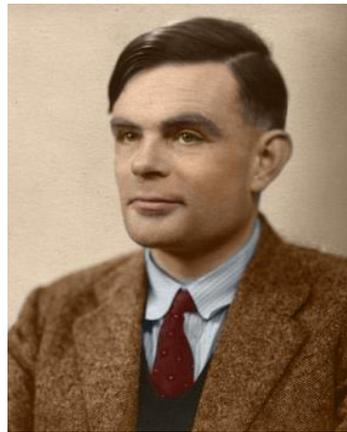
es decir, mediante modelos y algoritmos matemáticos; y c) Las computadoras representaría la herramienta técnica operativa adecuada para lograr la automatización o digitalización de la inteligencia humana. Para entonces la computadora sólo tenía una década de desarrollo (Mora y Wind, 2023a 2023b y otros/as autores/as). Se considera que este encuentro constituye precisamente el momento en el que se funda la Inteligencia Artificial, el cual cumplirá en poco tiempo setenta años, primero como campo de investigación cívico-militar y posteriormente como espacio de implementación comercial, económica y social. Si bien este grupo de inquietos/as investigadores/as no se mantuvo intacto durante el tiempo, la mayoría de ellos/as continuaron desarrollando, desde sus respectivas disciplinas científicas, la temática de la Inteligencia Artificial, logrando avances significativos una década después. Así, por ejemplo, Jaeger y Dacorogna (2024, 225) resumen esta interesantísima historia de la siguiente manera:

„La idea (teórica) de hacer que un ordenador funcione sobre la base (simplificada) de neuronas como clasificadores binarios basados en la estimulación externa se remonta al año 1943, cuando McCulloch y Pitts desarrollaron lo que hoy se conoce como „perceptrón“. En 1950, Alan Turing explica en su artículo „Computing Machinery and Intelligence“ cómo se pueden construir máquinas inteligentes y cómo se puede poner a prueba su inteligencia. Ya en 1951, Marvin Minsky construyó la primera máquina de aprendizaje cableada llamada SNARC. Cuatro años más tarde, se presentó el primer programa de IA en el Dartmouth Summer Research Project on Artificial Intelligence (DSPRAI), un taller de verano de ocho semanas de duración. Esta fue la base de la investigación en IA durante las siguientes décadas. También en esta conferencia de 1956, John McCarthy utilizó por primera vez el término „inteligencia artificial“. En 1958, Marvin Minsky y John McCarthy fundaron el MIT AI Group, que se convirtió en lo que hoy se conoce como el famoso Laboratorio de Ciencias Informáticas e Inteligencia Artificial del MIT. John McCarthy también fue el inventor de LISP en 1958, el segundo lenguaje de programación más antiguo adecuado para la IA. Él y sus socios dijeron, algo interesante para hoy en día: „La inteligencia artificial no es, por definición, una simulación de la inteligencia humana“. Más tarde, John McCarthy se trasladó a Stanford y fundó en 1963 el no menos famoso Laboratorio de Inteligencia Artificial de Stanford (SAIL)“.

En casi setenta años de evolución de la IA, la misma ha experimentado diversas etapas, unas de algarabía y entusiasmo, mientras que otras de disolución, cansancio, pausas y tal vez retroceso, son las llamadas etapas de inviernos de la IA. Por supuesto que, durante los últimos años, tal vez una década, ella ha tenido un avance vertiginoso, en forma ascendente, acelerada y exponencial, cubriendo buena parte de la actividad humana (autor sobre avance y retroceso de la IA). Sin embargo, todavía podemos decir que ningún sistema, tales como reconocimiento de voz, procesamiento de imágenes y aprendizaje profundo, etc. iguala a la versatilidad, habilidad, destreza y

potencialidad del cerebro humano. A pesar del gran desarrollo neurocientífico, el funcionamiento de nuestro cerebro sigue siendo un enigma, un misterioso órgano de células nerviosas intrínsecamente interconectadas, una complejidad de magnitud galáctica, tal vez imposible de igualar y mucho menos de superar; especialmente cuando hablamos de emociones, de creencias, de espiritualidad, de sentimientos, de creatividad compleja, de autonomía pensante y, por supuesto, de la potencialidad de las percepciones del mundo socionatural, lo cual nos permite la comprensión e interpretación de nosotros/as mismos/as, de los /as demás y del mundo de vida también complejo, todo lo cual difícilmente podría ser sustituido por una máquina, por mucho que ella sea superinteligente (Facundo Manes).

Figura 2: precursores de la IA.



a) Retrato de Gottfried Leibniz (1646-1716)

https://de.wikipedia.org/wiki/Gottfried_Wilhelm_Leibniz

b) Foto de Alan Turing

<https://static.wikia.nocookie.net/genios-computacion-1002/images/0/05/B6YvqbC.png>;

c) Modelo en madera del concepto de la Máquina de Alan Turing

<https://revistagalileu.globo.com/Tecnologia/noticia/2018/03/origem-da-computacao-maquina-de-turing-e-construida-em-madeira.html>

d) Modelo en metal del concepto de la Máquina de Alan Turing

<https://dramscome.blogspot.com/2021/06/original-alan-turing-machine-alan.html>

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica



a) Foto de Kurt Gödel (1906-1978) y Albert Einstein (1879-1955).

https://konyvultura.kello.hu/storage/upload_images/articles/article_BkIX0vJhDG2QYsC/articles_ozkiUSya13R9vqf_1705333946.png

b) Retrato de René Descartes (1596-1650).

<https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQDXgsSZNzOX6zkprnljHEQb8tEsS8uQAYWYA&s>

c) John Rogers Searle (1932-).

Encuentro académico y personal en la Universidad de Heidelberg.

Una de las citas más frecuentes (muchas veces sin referencias) realizadas en torno a la inmensa potencialidad del cerebro humano es del premio nobel de medicina Stanley Prusiner (2014, 248), quien señaló que: „La neurociencia es, con diferencia, la rama más apasionante de la ciencia, porque el cerebro es el objeto más fascinante del universo. Cada cerebro humano es diferente: el cerebro hace que cada persona sea única y define quién es“; mientras que el estudioso del cerebro humano Neil deGrasse Tyson (2017, 132), divulgador e investigador de temas relacionados con el cerebro humano, también señala que „Cada pensamiento, cada emoción, cada recuerdo que has tenido, es el resultado de 86 mil millones de neuronas comunicándose en patrones que ni los mejores supercomputadores pueden replicar“. Igualmente, Eric Kandel (2007), premio nobel de medicina en el 2000 por sus investigaciones sobre los mecanismos moleculares de la memoria considera que el cerebro humano sigue siendo único en su capacidad para compatibilizar e integrar aspectos biológicos, experienciales, emocionales y sensitivos, algo que la IA aún no puede replicar y, muy probablemente no lo podría hacer puesto que ella simplemente carece de vida, de sentimientos, de dolor y alegría naturales, etc.

Nadie se atrevería (salvo personas mentalmente enfermas) a maltratar a un ser querido, pero sí estaría dispuesto (tal como ocurre cada día) a construir, deconstruir y destruir máquinas, aún las más versátiles e inteligentes. Así sucesivamente, existe una gran cantidad de investigadores/as, filósofos, sociólogos, matemáticos, físicos/as y neurocientíficos, en general, que coinciden en torno a la complejidad del cerebro humano, el cual difícilmente podrían ser remplazado o superado por las máquinas, aún las más potentes e inteligentes que puedan ser construidas en el futuro. No se trata de ser pesimista, sino sencillamente de comparar un ser vivo que siente, sufre, se

alegra, tiene necesidades múltiples, primarias y secundarias, etc., con máquinas que sólo dependen de interconexiones puramente mecánicas, de algoritmos abstractos sin vida, basados en señales e informaciones electrónicas inertes. Por otra parte, el eminente científico, lingüista, analista y crítico político Chomsky (2023, s/p) ha señalado recientemente en un artículo publicado en el periódico New York Times, lo siguiente:

„Es a la vez cómico y trágico, como podría haber señalado Borges, que tanto dinero y tanta atención se concentren en una cosa tan pequeña, algo tan trivial cuando se contrasta con la mente humana, que a fuerza de lenguaje, en palabras de Wilhelm von Humboldt, puede hacer „un uso infinito de medios finitos“, creando ideas y teorías de alcance universal. La mente humana no es, como ChatGPT y sus similares, un pesado motor estadístico de comparación de patrones que se atiborra de cientos de terabytes de datos y extrapola la respuesta conversacional más probable o la respuesta más probable a una pregunta científica. Por el contrario, la mente humana es un sistema sorprendentemente eficiente e incluso elegante que funciona con pequeñas cantidades de información; no busca inferir correlaciones brutas entre puntos de datos, sino crear explicaciones“.

Las influencias de tres grandes matemáticos en Alan Turing

Antes de adentrarnos en el mundo de las suposiciones matemáticas, algorítmicas y tecnológicas de Alna Turing, es importante resaltar que, desde hace siglos, los seres humanos han estado preocupados por el pensamiento, la razón y el funcionamiento de la mente humana.

El mismo filósofo griego Aristóteles, dijo cuatro siglos antes de nuestra era, que probablemente habría una forma de comprender el pensamiento humano; habiéndose establecido cuatro a cinco siglos después un conjunto de reglas que ayudaría a comprender el pensamiento lógico de las personas. Y diecisiete siglos después de las apreciaciones de Aristóteles, Raimundus Lullus trabajó en el diseño de una máquina que tenía por finalidad obtener enunciados lógicos de la combinación puramente mecánica de ciertas palabras básicas, mientras que el gran matemático, físico y filósofo Gottfried Wilhelm Leibniz (a quien nos referiremos más adelante como una de las figuras más influyentes en el pensamiento de Alan Turing) trabajó varios años, concretamente en la conformación de su famosa máquina calculadora para las operaciones básicas aritméticas de suma, restar, multiplicar y dividir. Eso fue por allá en 1670.

Esta máquina la podemos apreciar en el museo de Gottfried Wilhelm Leibniz en la ciudad de Hannover (ver la figura 3). Cuatro siglos después, George Boole tuvo la peripecia intelectual de desarrollar la famosa lógica formal, las cuales fueron denominadas por él como reglas del pensamiento. La lógica

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

booleana es un concepto matemático de mucha aplicabilidad en la informática y digitalización modernas.

Figura 3: Máquina calculadora y el monumento a George Boole



a) Máquina calculadora de cuatro tipos, Gottfried Wilhelm Leibniz, construida en Hannover alrededor de 1690, Biblioteca Gottfried Wilhelm Leibniz, Hannover.

<https://www.uni-hannover.de/de/universitaet/profil/leibniz/leibnizausstellung#:~:text=Die%20einzig%20erhaltene%20Leibniz-Rechenmaschine,Wissenschaftskreisen%20und%20Fürstenhäusern%20des%2017.>

b) El monumento a George Boole se encuentra frente a la estación Lincoln, esta estatua conmemora la vida del matemático y académico local George Boole (1815-1865).

https://s1.geograph.org.uk/geophotos/07/55/14/7551421_b82f52b5_1024x1024.jpg.

Es decir, la idea de construir máquinas de pensar e intentarlo realmente en la práctica no es reciente, tiene realmente muchos siglos, Higuera (2018, 342 y 345) señala, por ejemplo, lo siguiente, haciendo referencia a algunos de los personajes que se han ocupado del tema, incluyendo por supuesto a Raimundus Lullus:

„La expresión „Máquina-de-pensar“ —*Denkmaschinen*— fue usada por Fritz Mauthner (Mauthner, 1910: 180) como un apelativo, un tanto «familiar», ya a finales del siglo xix, para equiparar el *Ars magna* luliano con la máquina lógica de Allan Marquand (Marquand, 1886: 303; Peirce, 1887: 166; Buck y Hunka, 1999: 22). Dicha expresión sirvió a Mauthner para proponer que el lenguaje, por medio de un conjunto de representaciones y procedimientos lógico-matemáticos, puede convertirse en un modelo —inherente y general— de las operaciones del pensamiento. A mediados del siglo xx Alan Turing (Turing, 1997: 29) afirmó que expresiones en las cuales se atribuye a una „máquina“ la diversidad operativa del pensamiento humano podrían sonar un tanto atrevidas, a pesar de que estas apuntan hacia un proyecto científico realizable en un tiempo indeterminado. Este tipo de afirmaciones han sido clasificadas por Searle (Searle, 1997: 183) dentro de las tres formas en que se puede formular la relación entre „máquinas“ y „mentes“... „Las

figuras del Arte luliano cumplen funciones de cifrado que son semejantes: los signos en los círculos corresponden a un cierto conjunto de significados (denominados *dignitas-principia*), el movimiento de estas figuras genera una serie de combinaciones, o fórmulas, de las que se extraen enunciados, y también hay tablas de resultados combinatorios en los que la alteración de un signo cambia la atribución del significado de la fórmula para que esta sea interpretada de otra manera (Ramon Llull, 1984: 196-206; 1986: 10-18; 2003: 9-24; 2007: 8-16)“.

Por supuesto que es durante el siglo pasado cuando los/as investigadores/as, especialmente en el campo de las matemáticas, avanzaron con mayor rapidez en el desarrollo de aparatos concretos que podrían constituirse en máquinas pensantes e inteligentes, tal como ocurrió durante los años veinte y treinta del siglo XX. Además de los dispositivos de cálculo mecánicos para facilitar el trabajo de oficina, tales como las máquinas de escribir (Mora y Wind, 2023a y 2023b), se utilizaron máquinas de tarjetas perforadas electromecánicas para tareas de procesamiento de datos de mayor magnitud. Parece ser que Martín Lebeis utilizó la idea del cerebro eléctrico a finales de la segunda década del siglo pasado (Müller, 1939). Aquí es muy importante mencionar las palabras de Dittmann y Seising (2020, 9), quienes señalan al respecto que:

„En el prefacio del libro de texto de 1939 sobre fotocopiadoras titulado *Máquinas controladas por sensores*, los autores afirmaron que se podía hablar de una máquina „pensante“. Estas máquinas escaneaban mecánicamente una pieza maestra, transferían el contorno de la superficie a una herramienta y, de esta forma, copiaban la forma. Hoy en día, difícilmente asociaríamos esta máquina con la inteligencia o el pensamiento, por lo que dichos términos deberían situarse en su contexto histórico. En la década de 1930, incluso se decía que los ascensores podían „casi pensar““.

Alan Turing (1912-1954), un inquieto joven matemático londinense trabajó en el campo de la lógica, informática, criptografía, inteligencia artificial y teoría de las probabilidades. Su trabajo fue pionero en el desarrollo de la informática moderna y, por supuesto, en el avance de lo que hoy conocemos como Inteligencia Artificial. Alan Turing creó la conjetura, prueba o experimento más conocido como Test de Turing, con lo cual trató de demostrar la idea incipiente de inteligencia artificial (todavía no se conocía el término) y el concepto de máquina de computación abstracta llamada máquina de Turing. En el campo de la biología teórica, su nombre se asocia al mecanismo de Turing de formación de estructuras bioquímicas. Es muy importante destacar que buena parte de sus trabajos escritos, aún en el campo de las probabilidades, no fueron publicados durante varias décadas, aún después de su muerte, ya que él trabajó para los servicios secretos ingleses durante el desarrollo la Segunda Gran Guerra Europea, contribuyendo de manera altamente significativa a descifrar los mensajes de los militares alemanes. El desarrollo posterior de la informática, la tecnología computacional y la inteligencia artificial se debe en

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

gran medida a sus aportes en torno a la concepción, configuración y experimentación de las máquinas calculadoras inteligentes.

Tres de sus promotores matemáticos a quienes Turing seguía y admiraba con pasión fueron René Descartes (1596-1650), quien, como matemático, filósofo y físico, creador de la geometría analítica y la filosofía moderna, se preocupaba no sólo por aspectos de las ciencias naturales y matemáticas, sino también por temas relacionados con la mente, el pensamiento y la conformación abstracta del lenguaje. Al buscar las raíces y génesis del debate en torno a la historia y desarrollo de la IA, nos encontramos obviamente con sus consideraciones en torno a la posibilidad de que pudiesen existir máquinas capaces de producir mensajes y acciones similares a las de los seres humanos. Él para entonces ya suponía la posibilidad de la conformación de máquinas que pudiesen producir palabras y actuar con las personas, aunque dudaba claramente de la sustitución del humano por las máquinas. Dejemos aquí al mismo Descartes (1637/2010, 79-80) que nos muestre, en la siguiente amplia cita, sus consideraciones en torno a las potenciales semejanzas entre el ser humano y la máquina:

„Y aquí me extendí particularmente haciendo ver que si hubiese máquinas tales que tuviesen los órganos y figura exterior de un mono o de otro animal cualquiera, desprovisto de razón, no habría medio alguno que nos permitiera conocer que no son en todo de igual naturaleza que esos animales; mientras que si las hubiera que semejasen a nuestros cuerpos e imitasen nuestras acciones, cuanto fuere moralmente posible, siempre tendríamos dos medios muy ciertos para reconocer que no por eso son hombres verdaderos; y es el primero, que nunca podrían hacer uso de palabras u otros signos, componiéndolos, como hacemos nosotros, para declarar nuestros pensamientos a los demás, pues si bien se puede concebir que una máquina esté de tal modo hecha que profiera palabras, y hasta que las profiera a propósito de acciones corporales que causen alguna alteración en sus órganos, como, verbi gratia, si se la toca en una parte, que pregunte lo que se quiere decirle, y si en otra, que grite que se le hace daño, y otras cosas por el mismo estilo, sin embargo, no se concibe que ordene en varios modos las palabras para contestar al sentido de todo lo que en su presencia se diga, como pueden hacerlo aun los más estúpidos de entre los hombres; y es el segundo que, aun cuando hicieran varias cosas tan bien y acaso mejor que ninguno de nosotros, no dejarían de fallar en otras, por donde se descubriría que no obran por conocimiento, sino sólo por la disposición de sus órganos, pues mientras que la razón es un instrumento universal, que puede servir en todas las coyunturas, esos órganos, en cambio, necesitan una particular disposición para cada acción particular: por donde sucede que es moralmente imposible que haya tantas y tan varias disposiciones en una máquina que puedan hacerla obrar en todas las ocurrencias de la vida de la manera como la razón nos hace obrar a nosotros“.

El segundo, Gottfried Leibniz (1646-1716), filósofo, matemático, jurista, historiador y consejero político de principios de la Ilustración que también influyó considerablemente en los trabajos e inspiraciones de Alan Turing y, por supuesto, al desarrollo actual de la IA, intentó diferenciar, a principios del siglo XVI, los conceptos de manipulación sintáctica y verdadera comprensión (semántica), reflexionando en torno a la posibilidad de que las máquinas podrían alcanzar una verdadera comprensión o más, aún si ellas podrían llegar a ser inteligentes (Leibniz, 1714).

En tercer lugar, Alan Turing siguió con admiración los logros de Kurt Gödel en cuanto a su consideración matemática (demostración realizada en 1931) que se refiere a la existencia de enunciados lógicos que son verdaderos, pero no necesariamente demostrables. Este matemático también diseñó un lenguaje universal para codificar cualquier proceso formalizable e, igualmente, conformó y demostró los teoremas de incompletitud, los cuales sostienen que en cualquier sistema formal suficientemente complejo hay enunciados que no se pueden decidir. El primer teorema de incompletitud señala que un sistema formal suficientemente potente (como la aritmética de Peano) es incompleto, es decir, existen proposiciones verdaderas dentro del sistema que no pueden ser demostradas usando sus propias reglas; mientras que el segundo teorema de incompletitud indica que ningún sistema formal consistente puede demostrar su propia consistencia. Es decir, si un sistema es lo suficientemente poderoso como para incluir la aritmética, no puede probar dentro de sí mismo que es libre de contradicciones. En vista de que la IA pretende, por lo menos a largo plazo, reproducir el pensamiento y las acciones mentales humanas, así como la resolución de complejos problemas con la ayuda de las máquinas, se requiere de sistemas formales y, especialmente, razonamientos lógicos basados en algoritmos. Aquí es donde entra a jugar un papel fundamental los aportes lógicos y matemáticos de Kurt Gödel, sentando las bases para comprender los alcances y límites teórico-prácticos del desarrollo actual futuro de la IA.

Tal como lo hemos mencionado en varias oportunidades (Wind y Mora, 2023a y 2023b), Alan Turing es ampliamente conocido por sus conjeturas relacionadas con las máquinas „inteligentes“, aunque el término conjetura no es exactamente de carácter matemático, como por ejemplo las conocidas conjeturas de Poincaré o de Goldbach. Algunos de sus supuestos son los siguientes:

La conjetura de Alana Turing sobre la Inteligencia Artificial. Él propuso en su artículo „Computing Machinery and Intelligence“ (Turing, 1950) que una máquina podría, en principio, imitar la inteligencia humana hasta el punto de no poder ser diferenciada de un ser humano al establecerse una conversación entre ambos con una tercera persona. Esta hipótesis dio origen al famoso *Test de Turing*, que sigue siendo un referente en la inteligencia artificial actual. La

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

conjetura de Turing sobre la Computabilidad. En su trabajo sobre las conocidas *máquinas de Turing*, propuso que cualquier problema matemático computable puede ser resuelto por una máquina de Turing. Esta idea es la base de la *Tesis de Church-Turing*, aunque técnicamente es una hipótesis y no un teorema formalmente demostrado.

La conjetura de Turing en Biología. Este importante matemático precursor de lo que hoy tenemos como „inteligencia artificial“ también propuso un importantísimo modelo matemático para explicar la morfogénesis, lo cual consiste en el proceso por el cual los organismos desarrollan patrones y formas en la naturaleza. Su famosa teoría sobre reacción-difusión sigue jugando un papel central en la biología matemática.

El Test de Turing o la prueba de la inteligencia de las máquinas

El famoso test propuesto por Alan Turing en 1950 plantea una hipotética conversación escrita entre un entrevistador humano con dos participantes, un segundo ser humano y una máquina computacional („inteligente“). Ahora bien, si el interrogador no puede distinguir de manera confiable cuál de los dos es la máquina, se concluye que ésta ha superado el test y puede considerarse como un aparato inteligente. Este criterio es de carácter esencialmente funcionalista; es decir, la conjetura de Alan Turing podría ser enunciada de la siguiente manera *sin una máquina digital pudiese imitar convenientemente el comportamiento y el razonamiento de un ser humano, entonces podríamos considerar o afirmar que dicha máquina es en efecto inteligente.*

Muchas veces se menciona a Alan Turing por su aporte en cuanto a descifrar el famoso código enigmático durante el desarrollo de la Segunda Gran Guerra Europea; sin embargo, él dedicó gran parte de su corta vida a investigar en torno a las potenciales relaciones entre el pensamiento y las máquinas, entre el cerebro y la electrónica. Su famoso trabajo, publicado en 1950, trata precisamente sobre la interrogante ¿Pueden pensar las máquinas? (Turing, 1950). Allí es donde él propone realmente su legendaria prueba empírica o el también conocido *Test de Turing*.

Es aquí cuando tiene lugar el inicio de la estrecha relación entre la inteligencia artificial y la conciencia humana, lo cual ha motivado la controversia científica, pero también popular, entre muchos/as investigadores/as en torno a las posibilidades de que las máquinas iguallen a los seres humanos en cuanto al pensamiento y el razonamiento (Villagómez, 2025; Lenzen, 2020 y 2023).

Si bien durante los últimos años se ha producido una gran producción de artículos científicos, por un lado, y opiniones populares, por el otro, tal vez muchas de ellas son producto del uso de los mismos sistemas algorítmicos ganadores de datos e informaciones. Pero el debate en torno a esta temática

es mucho más antigua, tiene por supuesto su propia historia, tal como lo señalan resumidamente Dittmann y Seising (2020). En tal sentido, queremos ampliar un poco más, en este apartado, parte de esa historia, tomando en cuenta algunos/as de los/as autores/as que ya desde hace bastante tiempo han reflexionado sobre la temática (Shannon, 1948, 1949 y 1950; Goldschlager y Lister, 1986; McCorduck, 1987; Dreyfus y Dreyfus, 1987; Capron y Reed, 1990; Dewdney, 1995; Rechenberg y Pomberger, 2002; Hellige, 2004; Lungarella; Iida; Bongard y Pfeifer, 2007; Russell y Norvig, 2012; Lankau, 2017a y 2017b; Seipold, 2018; Witt y Gloerfeld, 2018; Herbert, 2018; Strauß, 2019; Hess, 2022; Mora y Wind, 2023a y 2023b; entre muchos/as otros/as).

En este sentido, los seres humanos, especialmente quienes se han interesado por las matemáticas, siempre han estado interesados/as por establecer mecanismos para la resolución automática de problemas, especialmente en el campo de la teoría de números; sin embargo, es a partir del inicio de la segunda mitad del siglo pasado (exactamente hace 75 años), que la investigación algorítmica se centró fundamentalmente en estudiar potenciales habilidades cognitivas de los seres humanos y cómo ello podría ser simulado o también realizado por las máquinas. Ya hemos señalado en párrafos anteriores del presente documento que los participantes de la Conferencia de Dartmouth tuvieron la habilidad intelectual y práctica, en 1957, de desarrollar el famoso Solucionador General de Problemas, el cual consistía en un programa computacional que pretendía (y lo hacía en cierta forma) imitar el razonamiento humano.

De manera simultánea o paralela, el MIT trabajó ampliamente en los denominados micromundos, lo cual tenía que ver con dividir el todo en partes; así por ejemplo la semejanza de un ser humano mediante un robot, se podría estudiar más fácilmente, haciendo simulaciones de brazos, piernas, etc. Estas partes o micromundos podían hacer ciertas actividades sencillas, pero obviamente no lograban resolver problemas de mayor complejidad, menos simular el pensamiento y, muchos menos, razonar, tal como podría ocurrir, según los más optimistas de la IA, en un futuro cercano. Poco a poco, los/as matemáticos/as e informáticos/as, entre otros/as científicos/as, lograron resolver muchos problemas técnicos y avanzar en el campo de las potencialidades computacionales.

Este desarrollo continuado estuvo estrechamente ligado al avance de los lenguajes computacionales, tales como LISP, desarrollado en la década de los cincuenta del siglo pasado, el cual fue elaborado con la finalidad de manipular listas grandes de datos e informaciones, convirtiéndose en uno de los más utilizados para el posterior desarrollo de la IA. Dos décadas después, cerca de los años setenta, también en el siglo XX, tuvo lugar en Francia un importante progreso de estos lenguajes computacional, con la conformación de PROLOG,

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

el cual estuvo basado en la lógica formal booleana, la cual es la base de los sistemas expertos informáticos. El lenguaje LOGO, trabajado particularmente en la década de los sesenta, y que fue una continuidad del lenguaje LISP, tenía por objetivo enseñarles en la escuela programación a los/as niños/as. Estas iniciativas no tuvieron mucho éxito por su complejidad y la carencia de estrategias didácticas apropiadas para el tratamiento de la computación en la escuela, problemática que aún sigue presente en la actualidad. Todo estos experimentos computacionales e informáticos pusieron la piedra fundamental de la actual IA simbólica, por un lado, pero también permitieron mostrar las grandes limitaciones, para entonces, de la simulación de la inteligencia humana, en términos de la máquina de Turing, lográndose demostrar que la inteligencia humana no podía reducirse a reglas predefinidas. Esta dificultad le permitió a los/as matemáticos/as e informáticos/as emprender nuevos caminos para mejorar considerablemente los algoritmos que predominan actualmente en el mundo complejo de la IA.

El segundo gran adelanto de la era computacional de la segunda mitad del siglo pasado consistió en la conformación de sistemas basados en el procesamiento de conocimientos, particularmente en el mundo militar, comercial, financiero y científico. Por ello, la población en general no tuvo tanto acceso a tales beneficios algorítmicos. Las primeras investigaciones en IA lograron éxitos importantes en el adelanto de *sistemas expertos*, siendo uno de los principales pioneros el DENTRAL, creado por Edward Feigenbaum en Stanford en 1969. Dicho sistema, muy avanzado y novedoso para el momento histórico, podía determinar con márgenes de error muy pequeños la estructura molecular de compuestos orgánicos a partir de datos espectrométricos, demostrando con ello el alto potencial práctico de la IA.

Un segundo perfeccionamiento de esta tecnología consistió en el invento del MYCIN, el cual consistía en un sistema experto que permitía diagnosticar infecciones sanguíneas. Este programa computacional ha servido como un referente altamente significativo para la incorporación exitosa posterior de la IA en el campo de la medicina. El mismo tenía la capacidad de funcionar con 450 reglas prácticas, las cuales fueron construidas a partir de entrevistas a personal médico y paramédico, con lo cual se podía tener una importante correlación de certeza en los resultados realizados a los diagnósticos sanguíneos. Al hacer algunas pruebas comparativas, se pudo constatar el alto nivel de equivalencia con las respuestas suministradas por los mismos seres humanos, en su mayoría médicos y paramédicos. A pesar de estos excelentes resultados, el programa fue decayendo en la medida que aumentaba, por otro lado, el nivel de incertidumbre probabilístico. Para la década de los ochenta del siglo XX, estos sistemas, basados en su mayoría en el procesamiento de conocimientos expertos, alcanzaron una importante madurez técnica, lo cual permitió su ofrecimiento comercial a un número mucho mayor de personas en el ámbito internacional, hablándose para

entonces de la quinta generación de computadoras, a las cuales podrían tener acceso buena parte de la población mundial particularmente los/as investigadores/as, docentes, empresarios y comerciales.

Países asiáticos como Japón y China, por ejemplo, invirtieron grandes cantidades de dinero y talento humano (alrededor de los primeros años de la década de los ochenta) en el desarrollo de esta tecnología, logrando resultados altamente sorprendentes. Uno de los objetivos consistía en integrar sistemas expertos en computadoras personales, lográndose una interacción más fluida entre el lenguaje de máquina y el lenguaje natural. Por supuesto que durante las últimas tres décadas del siglo pasado, tuvo lugar una significativa competencia entre muchos países por el dominio de los adelantos sobre los sistemas computacionales de procesamiento de datos e informaciones. Así, por ejemplo, Jaeger y Dacorogna (2024, 83) señalan que:

„El mayor desarrollo de la IA es de gran importancia en la política social y global: el país con la IA más avanzada probablemente se convertirá en la potencia económica y militar dominante en este planeta. Actualmente, dos países compiten por la supremacía mundial en IA: Estados Unidos y China. China ha logrado avances significativos en los últimos años y ahora incluso aspira al primer puesto. Los europeos hace tiempo que se quedaron atrás en esta carrera y se vieron relegados a la categoría de extras. La base del liderazgo de Estados Unidos y China no son investigadores más inteligentes, mejores algoritmos de IA o mejores programadores informáticos, sino simplemente la disponibilidad de datos. En este contexto también se les considera „el petróleo del siglo XXI““.

El avance de la computación funcional también la podemos encontrar en el interés que han tenido los/as matemáticos/as e informáticos/as en hacer simulaciones y desarrollar estrategias lúdicas, especialmente en el área de los juegos de mesa. Es decir, el desarrollo histórico de la IA también ha estado asociado a la realización de experimentos con muchos juegos de mesa ampliamente practicados en muchas culturas.

Ellos se han constituido como un banco de pruebas ideal para la IA, puesto que los mismos contienen reglas muy claras, fases y procedimientos finitos, así como entornos altamente controlados. Norbert Wiener (1947) predijo hace más de 75 años que en el futuro habría alguna máquina lo suficientemente inteligente que podría jugar y ganar a los seres humanos mediante un proceso basado en multiplicidades de movimientos mecánicos. Y es precisamente en 1997, medio siglo después, que la Deep Blue venció al campeón mundial de ajedrez Gari Kaspárov (Wind y Mora, 2023a). En la actualidad la IA ha avanzado considerablemente en esta tarea, de tal manera que los/as mejores/as jugadores/as de este significativo e inteligente juego tienen grandes dificultades para obtener victorias. El segundo juego de mesa, también muy complejo, es el Go. Hace casi 10 años el famoso programa

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

AlphaGo derrotó al ampliamente conocido jugador Lee Sedol; en la actualidad muchos/as jugadores/as también se enfrentan con dificultad a las máquinas inteligentes en torno a este legendario e importante juego.

El programa Watson, construido por IBM en el 2011 para jugar Jeopardy, le ganó simultáneamente a grandes campeones humanos, interpretando preguntas con juegos de palabras, demostrando avances en procesamiento de lenguaje natural, lo cual ha contribuido considerablemente al desarrollo de programas de traducción, los cuales han adquirido durante los dos últimos años una perfección inigualable, en cuanto a velocidad y cantidad de textos traducidos, por la capacidad natural del ser humano. El famoso programa Libratus ha vencido, ya en el 2017, a profesionales expertos en el juego Póker, usando estrategias adaptativas, incluso con información incompleta. Quienes apoyan ciegamente a la IA, consideran que estas máquinas han usado una inteligencia igual o superior a la humana, en términos de pensamiento, mientras que aquéllos que aún tenemos ciertas dudas en cuanto a la capacidad mental de la IA, consideramos que se trata todavía de procesos complejos de simulación de inteligencia.

La Habitación China y la primera gran crítica filosófica al Test de Turing y a la IA

Desde que en el 2023 se puso en boga el uso masivo de la llamada Inteligencia Artificial, vemos igualmente el surgimiento del debate internacional en torno el funcionamiento de esta tecnología en cuanto a la igualdad o superación de la inteligencia humana, especialmente en el campo de la producción de saberes conocimientos en los diversos campos de las ciencias e investigación, pero también en las variadas manifestaciones del pensamiento semántico de los seres humanos. Este debate no es reciente, por supuesto, desde que John Searle publicó su famoso artículo „mentes, cerebros y programas“ (Searle, 1980), con frecuencia ha tenido lugar en diversos países y lenguas una discusión acalorada, vista desde diversas perspectivas, que aún no tienen consenso o no se aceptan los argumentos de la contraparte. Esta disputa gira en torno al famoso argumento de la habitación china, descrito por Johan Searle (1980, 1986 y 1995), quien ha elaborado diversas variantes en la medida que surgían posicionamientos teóricos (y también prácticos) contrarios a su experimento hipotético. En 1986, Johan Searle escribe, por ejemplo, todo un capítulo de su libro „Mente, cerebro y ciencia“, cuyo título es „¿Pueden pensar las máquinas?“

John Searle consideró, al inicio de la década de los ochenta, que era necesario distinguir dos tipos de inteligencia artificial para poder analizar y criticarla, en su caso particular, mediante un experimento mental denominado por él como la „habitación china“. El primer tipo lo denominó como Inteligencia Artificial débil, mientras que el segundo lo llamó Inteligencia Artificial fuerte. El

primero se refiere a sistemas que sólo permiten hacer imitaciones inteligentes, tales como reconocimiento de voz, traducciones, respuestas basadas en búsquedas, respuestas centradas en el cruce inteligente de grandes cantidades de datos e informaciones, diagnósticos médicos, etc. Ella constituye herramientas avanzadas, pero carece de conciencia. En el segundo caso, nos encontramos con máquinas que podrían tener estados cognitivos reales, equivalentes a la mente humana, lo cual sigue estando muy lejos de ser posible, teniendo John Searle todavía razón en sus conjeturas. Mientras la IA actual opera como una herramienta sofisticada de procesamiento de datos e informaciones, la búsqueda de una inteligencia artificial consciente continúa, inspirándose en modelos biológicos como las complejas redes neuronales humanas. Searle sigue pensando, casi cincuenta años después de su experimento mental de la Habitación China que ello sería imposible de lograr (Searle, 1980, 1986 y 1995).

La idea fundamental de John Searle (1980, 1986 y 1995) consiste en intentar demostrar que la respuesta a esta interrogante debe ser respondida en sentido estrictamente negativo. Es decir, no será posible que las máquinas puedan pensar, tal como lo hacemos los seres humanos, puesto que las máquinas por muy avanzadas o supuestamente inteligentes que pueden llegar a ser, ellas no podrán tener estados mentales complejos como los tenemos nosotros, las personas de carne y hueso, de células vivas.

Como respuesta a las inquietudes que tenía los/as informáticos/as, matemáticos/as, programadores/as y demás científicos/as, también en el campo de las ciencias sociales y humanísticas, John Searle (1980, 1986 y 1975) desarrolla su famoso experimento mental, cuyo objetivo era demostrar que las computadoras no entienden ni llegarán a entender realmente la esencia de los comportamientos humanos y su capacidad para la producción originaria de saberes y conocimientos, sino que ellas simplemente pueden reproducir, eso sí de manera muy bien ordenada, lo que han almacenado o recopilado mediante su alta capacidad de almacenamiento y procesamiento informático.

Para ello, John Searle (1980, 338 y ss.) propone un experimento mental para argumentar que la manipulación sintáctica de símbolos no es suficiente para generar comprensión o conciencia, en respuesta a posturas de la inteligencia artificial fuerte, que sostienen que una computadora puede „entender“ o „pensar“ si está debidamente programada. En el experimento, Searle (1980 y 1986) pide que imaginemos una situación hipotética en la que él es encerrado en una habitación sin ningún conocimiento del idioma chino. Dentro de esta habitación recibe tres lotes de caracteres chinos, sin entender su significado, junto con un conjunto de instrucciones escritas en inglés. Estas instrucciones le dicen cómo relacionar los diferentes grupos de caracteres, de manera puramente formal (es decir, solo por la forma de los símbolos, no por su

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

contenido). A pesar de no comprender el idioma, siguiendo meticulosamente las instrucciones, la persona que se encuentra en la habitación, completamente cerrada, logra producir respuestas en chino que, desde el exterior, parecen perfectamente naturales, ciertas, coherentes y con sentido, como si fueran generadas por un hablante nativo; es decir, por una persona cuyo idioma materno sea el chino. A ojos de un observador externo, dicha persona „responde en chino“ de forma indistinguible de una persona que verdaderamente lo domina.

Sin embargo, Searle (1980 y 1986) insiste en que, aunque el sistema pueda parecer que entiende el idioma, él mismo no comprende absolutamente nada de lo que está procesando. Sólo manipula símbolos sin saber qué significan. Por tanto, el hecho de que una máquina (o una persona siguiendo un programa o un conjunto de indicaciones bien precisas) pueda generar respuestas coherentes no implica que haya comprensión real, que sepa realmente lo que está haciendo, que tenga un sentido esencial, crítico, social y humano de lo que está produciendo. Searle (1980) concluye que las computadoras u ordenadores funcionan de manera análoga a su situación en la habitación: procesan datos según reglas sintácticas, pero sin comprensión semántica. Por eso, aunque una computadora u ordenador pueda simular informáticamente el lenguaje humano, esto no significa que entienda lo que dice, mucho menos, lo que realmente está haciendo en el marco y contextos en torno a los cuales se producen tales saberes y conocimientos. Así, este autor, hace cuatro décadas y media, critica la noción de que un programa de computadora, por muy avanzado y complejo que sea, pueda por sí solo generar mente y comprensión.

Es importante resaltar que este experimento mental de John Searle (1980, 1986 y 1995) argumenta claramente que tales programas, por muy justificados que sean no entenderán nunca las historias, los contextos, las realidades e interacciones entre multiplicidad de sujetos que estarían directa e indirectamente relacionados con tales saberes y conocimientos. Las respuestas que dictan están bien estructuradas y organizadas con base en las grandes cantidades de datos e informaciones acumuladas y disponibles, en su mayoría producto de la creatividad y producción humanas. Los computadores y sus respectivos programas no podrán, según la opinión de John Searle (1980 y 1986), comprender ningún lenguaje ni tener ningún otro estado mental, puesto que el lenguaje, el saber, el conocimiento y la comprensión están asociados intrínsecamente a la cultura, a la historia a las relaciones entre sujetos vivientes y pensantes. La razón de esta incapacidad de la llamada inteligencia artificial radica esencialmente en el hecho de que los programas informáticos estándar sólo realizan manipulaciones sintácticas o formales de símbolos.

Por supuesto que John Searle trata de argumentar, en contra de las máquinas supuestamente inteligentes, que las mentes humanas están constituidas biológicamente por células, proteínas, redes neuronales vivas que reciben, envían y producen conocimientos gracias a su complejidad neuronal natural, pero se olvida agregar que tal elaboración de saberes y conocimientos es posible sólo mediante la existencia de la complejidad del resto de los órganos o demás componentes del cuerpo humano, la totalidad de sus sentidos, los cuales nos permiten la interacción con el medio, con la realidad.

El cerebro humano en sí mismo y aislado no sería obviamente suficiente para pensar, crear, discernir y generar el saber y el conocimiento que lo caracteriza como ser creativo, crítico y discursivo. Un cerebro biológico aislado, por sí solo, no podría pensar lo suficientemente, tal como lo hace realmente en su conexión e interacción con la totalidad del cuerpo. Si, por el contrario, él está desconectado del cuerpo y de los órganos sensoriales, ya no tendría acceso a la realidad y, por tanto, tampoco se diferenciaría mucho de las potentes máquinas inteligentes actuales. Lo interesante, entonces, del ser humano, de la mente humana y de la creación de saberes y conocimientos es la existencia también de la estructura corporal sensorial de los humanos.

Antecedentes históricos de la Habitación China y la IA

Si bien el argumento de la Habitación China, propuesto por John Searle en 1980, continúa siendo, 45 años después, un hito relevante en la discusión sobre los límites de la llamada „Inteligencia Artificial“ (IA), su desarrollo se inscribe en una larga tradición filosófica, matemática y computacional, cuyas raíces preceden ampliamente a la explosión contemporánea de los sistemas de IA generativa. A continuación, presentamos un breve recorrido por algunos antecedentes significativos (Wind y Mora, 2023a; 2024).

Desde el siglo XVII, pensadores como René Descartes comenzaron a explorar la relación entre cuerpo, mente y razón, inaugurando una visión dualista que distinguiría entre la *res cogitans* (mente) y la *res extensa* (materia) (Descartes, 1641). En el siglo XVIII, filósofos ilustrados como David Hume e Immanuel Kant profundizaron en el problema del conocimiento y la mente humana, sentando las bases del idealismo trascendental y el empirismo, que posteriormente influirían en los debates sobre inteligencia y conciencia, los cuales aún siguen teniendo su importancia cuando hablamos del desarrollo histórico, actualidad y futuro de la IA.

Más adelante, Leibniz anticipó una lógica universal de pensamiento mediante símbolos, al imaginar una *lingua characteristic universalis* que permitiría resolver cualquier problema intelectual mediante cálculo mecánico (Leibniz, 1666). Este matemático alemán ha influido considerablemente en el campo de las matemáticas, al igual que René Descartes, especialmente en el área del

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

cálculo infinitesimal e integral. Es decir, el desarrollo de la IA ha estado asociado directa e indirectamente con las matemáticas, la física y la filosofía.

En el siglo XIX, George Boole desarrolló el álgebra binaria que posteriormente se convertiría en la base de la lógica computacional moderna (Boole, 1854). Ya en el siglo XX, Norbert Wiener propuso la cibernética como un modelo de control y comunicación en máquinas y organismos vivos, enfatizando la retroalimentación como principio operativo común entre humanos y artefactos (Wiener, 1948). Un hito decisivo fue la propuesta del Test de Turing, en 1950, donde Alan Turing sugirió que la inteligencia artificial podría evaluarse no por su comprensión interna, sino por su comportamiento exterior indistinguible del humano en una conversación (Turing, 1950). Esto dio paso a una visión funcionalista del pensamiento artificial, ampliamente difundida durante las décadas siguientes.

Sin embargo, varias voces comenzaron a problematizar esta perspectiva. En 1961, el filósofo soviético Anatoly Dneprov, en su cuento *El experimento*, ilustró cómo una computadora podía manipular símbolos sin entender su significado, anticipando la crítica central de Searle. Posteriormente, Lawrence Davis propuso en 1974 el experimento mental de la „habitación telegráfica“, donde una persona aislada podía responder mensajes en un idioma desconocido usando un manual, sin comprenderlos. En 1978, Ned Block desarrolló la analogía de la nación china, en la que millones de personas actuaban como una red simbólica de procesamiento sin comprensión real (Block, 1978).

A estos experimentos se suman las críticas fenomenológicas de Hubert Dreyfus, quien desde los años sesenta del siglo pasado denunció la incapacidad de la IA simbólica para emular el sentido común humano, debido a su desconexión con el cuerpo, el mundo y la experiencia encarnada (Dreyfus, 1972). En paralelo, Jerry Fodor propuso la hipótesis del lenguaje del pensamiento, según la cual el pensamiento humano podría estar codificado simbólicamente, un planteamiento que, si bien buscaba fundamentar la IA, fue también criticado por su reduccionismo (Fodor, 1975). Finalmente, John Haugeland acuñó el término GOFAI (Good Old-Fashioned Artificial Intelligence) para referirse de forma crítica a los modelos clásicos de IA basados exclusivamente en reglas lógicas (Haugeland, 1985).

Todas estas aportaciones configuran un trasfondo teórico e histórico que permite entender con mayor profundidad el sentido del argumento de Searle. Su crítica no se limita a una objeción técnica, sino que constituye una defensa de la dimensión semántica, intencional y experiencial del pensamiento humano, la cual no puede ser reducida a mera manipulación sintáctica de símbolos.

Diversas corrientes filosóficas contemporáneas sostienen que la conciencia es un fenómeno exclusivamente biológico, propio de los seres vivos, y que ninguna máquina —por sofisticada que sea— puede replicar genuinamente la experiencia consciente humana. Desde esta perspectiva, la conciencia no es solo el producto de una manipulación simbólica o computacional, sino el resultado emergente de complejos procesos neurobiológicos, afectivos y experienciales que aún no pueden ser reproducidos artificialmente (Damasio, 1999; Searle, 1992).

En este marco, el filósofo John Searle realiza una crítica contundente a los postulados de la inteligencia artificial fuerte, especialmente a la idea de que un sistema computacional pueda entender del mismo modo que lo hace un ser humano. Según Searle, los estados mentales no son meras estructuras formales; poseen un contenido semántico que trasciende su forma sintáctica. En sus propias palabras: „La sintaxis no es suficiente para la semántica, y los ordenadores, en tanto sistemas puramente sintácticos, carecen de semántica intrínseca“ (Searle, 1980, 417).

Esta tesis se articula magistralmente con el mencionado experimento mental de la Habitación China, partiendo del hecho que hipotéticamente un sujeto manipula símbolos de un lenguaje siguiendo un conjunto de instrucciones predeterminadas (un „programa“). Es importante reiterar que aunque logra generar respuestas que resultan indistinguibles de las de un hablante nativo, el sujeto carece por completo de comprensión sobre el significado de los símbolos que maneja. Esto demuestra que la ejecución de una manipulación sintáctica de símbolos no implica necesariamente la existencia de una comprensión semántica real. Searle utiliza esta analogía para refutar la afirmación de que las máquinas puedan entender simplemente por ejecutar programas. A diferencia de los humanos, que dotan de significado a sus pensamientos mediante experiencias vividas, corporalidad e intencionalidad, los ordenadores operan exclusivamente mediante reglas formales —un procesamiento sintáctico que no puede generar por sí mismo contenido semántico (Searle, 1992; Preston y Bishop, 2002).

Incluso figuras como Albert Einstein valoraban el poder de los experimentos mentales para explorar los límites del conocimiento humano. Él mismo describía su mente como un laboratorio para imaginar situaciones físicas imposibles de reproducir empíricamente (Holton, 1971). En esa tradición, el experimento de la Habitación China se presenta como una poderosa herramienta filosófica para cuestionar los límites del pensamiento computacional. Posteriormente, esta idea sería retomada y popularizada por el físico y matemático Roger Penrose, quien también argumenta que la mente humana no puede ser completamente reducida a algoritmos formales (Penrose, 1989).

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

En definitiva, la propuesta de Searle no niega que las máquinas puedan ejecutar tareas de forma eficiente; lo que objeta es que esa eficiencia no implica comprensión o conciencia. Como él mismo afirma, una máquina puede hacer como si comprendiera, pero esto es muy distinto a la comprensión real, que implica intencionalidad, experiencia y significado.

El surgimiento de las redes neuronales y el futuro de la IA

Los fundamentos de las redes neuronales artificiales se remontan a 1943, cuando el psicólogo Warren S. McCulloch y el matemático Walter Pitts identificaron los principios lógicos detrás de la actividad neuronal. Sus trabajos pioneros demostraron cómo las conexiones entre neuronas artificiales podrían fortalecerse mediante el uso conjunto de las mismas; tratándose de un concepto matemático-informático conocido como la Regla de Hebb, en honor al psicólogo y fisiólogo Donald Hebb (2002/1949). Este mecanismo de aprendizaje adaptativo sigue siendo la base del entrenamiento de redes neuronales modernas.

En la actualidad, estas redes se han convertido en herramientas indispensables para tratar y analizar los ampliamente conocidos Big Data, detectando patrones complejos en imágenes, sonidos y textos que escapan al análisis manual y visual del ser humano. Su capacidad para procesar información masiva las ha posicionado como pilares fundamentales del aprendizaje automático del mundo actual. Para corroborar esta consideración neuronal, podemos recordar mediante la siguiente cita (Kickuth, 2018, 111), los aportes iniciales de Santiago Ramón y Cajal y Donald Hebb:

„Ya en 1894, el español Santiago Ramón y Cajal postuló que la memoria podía formarse fortaleciendo la conexión entre neuronas existentes. En 1949, Donald Hebb, un fisiólogo canadiense, formuló la primera regla del aprendizaje: las neuronas simultáneamente activas fortalecen las conexiones; en inglés hay un dicho pegadizo: „lo que se enciende junto, se conecta junto“. Para explicarlo con más detalle: cuando una célula nerviosa A excita de forma permanente y repetida a una célula nerviosa B, la sinapsis se modifica de tal manera que la transmisión de señales se vuelve más eficiente. Esto aumenta el potencial de membrana en la neurona receptora“.

3. Breve descripción sobre el funcionamiento de la estructura de una red neuronal artificial

En los apartados anteriores hemos indicado, en varias ocasiones, que la IA consiste en un campo de la informática que busca incesantemente que las máquinas puedan „aprender“, „razonar“ y „tomar decisiones“, de forma parecida a los seres humanos. Uno de los modelos más utilizados para lograr

esto son las redes neuronales artificiales (RNA), lo cual se basa en el funcionamiento de las redes neuronales biológicas humanas (Lämmel y Cleve, 2012 y 2023 Furbach, Kitzelmann, Michaeliy y Schmid, 2024).

Las neuronas se caracterizan por dos tipos de extensiones de la membrana celular. Las dendritas reciben señales de otras neuronas y las transmiten al cuerpo celular. Cuando una neurona se activa, genera un potencial de acción, una señal eléctrica que viaja a lo largo del axón. Los potenciales de acción permiten la transmisión de la información a otras neuronas a través del axón. En muchas neuronas, el axón está rodeado por una vaina de mielina, que actúa como aislante y aumenta la velocidad de transmisión de la señal.

Modelo de la neurona biológica

Tal como lo explican Christoph von der Malsburg (1973) y Wolf Singer (1999), el modelo de neurona biológica es una representación conceptual y funcional del comportamiento de las células nerviosas del cerebro, conocidas como neuronas. Estas células son las unidades básicas del sistema nervioso y están especializadas en la recepción, procesamiento y transmisión de información mediante señales electroquímicas. Cada neurona se compone de tres partes principales: el soma (cuerpo celular), donde se integra la información; las dendritas, que reciben señales de otras neuronas; y el axón, que transmite impulsos hacia otras células. La comunicación entre neuronas se realiza a través de sinapsis, donde se liberan neurotransmisores que activan o inhiben la neurona receptora.

Desde una perspectiva funcional, la neurona biológica puede interpretarse como un sistema de entrada-procesamiento-salida. Las dendritas reciben múltiples señales sinápticas que son integradas en el soma. Si la suma de estas señales supera un cierto umbral de excitación, la neurona genera un potencial de acción que se propaga a lo largo del axón hasta alcanzar otras neuronas o células efectoras. Este mecanismo permite que las neuronas actúen como procesadores de información, capaces de tomar decisiones simples (como disparar o no disparar) basadas en la información recibida, lo que constituye la base de funciones cognitivas más complejas como el aprendizaje, la memoria o la percepción.

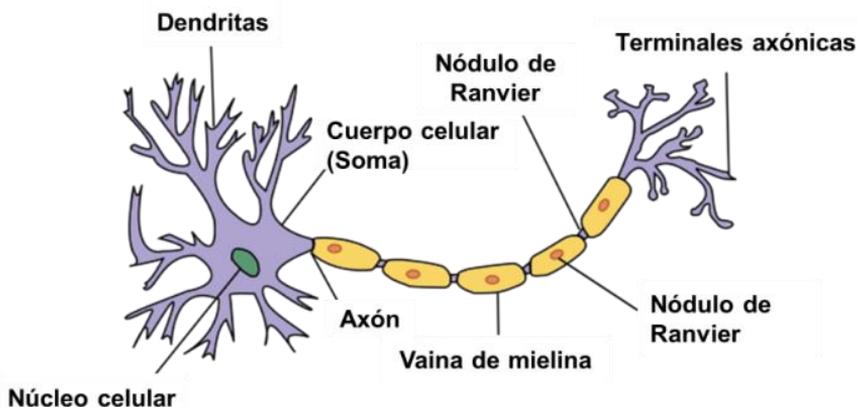
Este modelo biológico ha servido como inspiración para el desarrollo de las redes neuronales artificiales en el campo de la inteligencia artificial y la computación. En estos modelos computacionales, cada „neurona artificial“ simula el comportamiento de una neurona biológica mediante la recepción de entradas ponderadas, la aplicación de una función de activación (equivalente al umbral biológico) y la producción de una salida. Aunque simplificado (figura 4), el modelo de neurona biológica permite comprender los principios básicos

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

del procesamiento neuronal, y ha sido clave para vincular la neurociencia con disciplinas como la ingeniería, la psicología cognitiva y la informática.

Figura 4: Modelo de la neurona biológica

Dendritas



Modelo de neurona biológica

Fuente:

https://upload.wikimedia.org/wikipedia/commons/thumb/e/ec/Neuron_%28deutsch%29-1.svg/600px-Neuron_%28deutsch%29-1.svg.png.

El modelo de la neurona artificial

De la misma manera, existe una gran cantidad de autores/as, muchos/as de ellos/as recientes, quienes explican y analizan, con base en el modelo de la neurona biológica, el modelo de la neurona artificial; así, por ejemplo, Montavon, Samek y Müller (2018) y Montavon, Samek y Müller (2018) y Müller, Samek, Montavon, Vedaldi y Hansen (2019), el modelo de neurona artificial es una abstracción matemática inspirada en el funcionamiento de la neurona biológica.

Fue originalmente propuesto por Warren McCulloch y Walter Pitts en 1943, y ha evolucionado hasta convertirse en la unidad fundamental de las redes neuronales artificiales utilizadas en inteligencia artificial (IA). Su estructura simula el procesamiento de señales: recibe un conjunto de entradas (inputs), cada una con un peso asociado, realiza una suma ponderada de esas entradas y luego aplica una función de activación que determina la salida de la neurona. Esta salida puede servir como entrada para otras neuronas en una red.

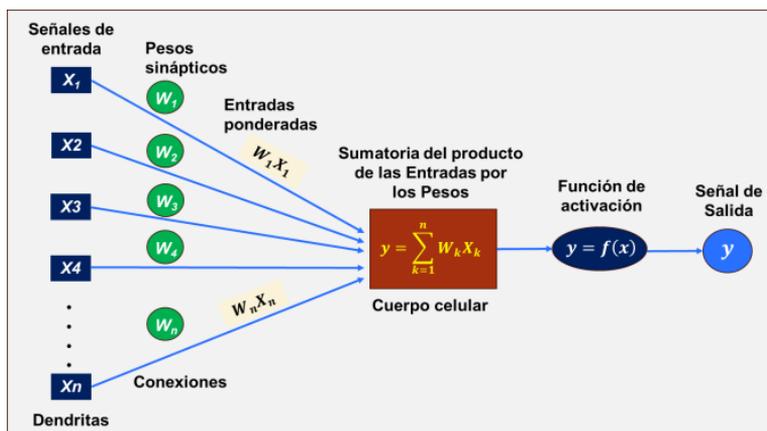
Funcionalmente, una neurona artificial opera mediante la fórmula:

$$y = f \left(\sum_{n=1}^n W_i X_i + b \right)$$

Donde x_i son las entradas y w_i son los pesos sinápticos, b el sesgo (bias), y f es una función de activación (como la sigmoide, ReLU o tangente hiperbólica). La función de activación introduce no linealidad, permitiendo que las redes neuronales aprendan relaciones complejas entre los datos de entrada y las salidas deseadas. Este modelo es la base de algoritmos de aprendizaje supervisado, como el perceptrón y el backpropagation, y también sustenta arquitecturas modernas como las redes convolucionales y las redes recurrentes.

El modelo de neurona artificial, mostrado en la figura 5, permite a los sistemas computacionales „aprender“ a partir de ejemplos, ajustando los pesos mediante algoritmos de entrenamiento para minimizar el error de predicción. Aunque es una simplificación del comportamiento neuronal real, su efectividad en tareas como reconocimiento de patrones, procesamiento del lenguaje natural y visión artificial ha demostrado su utilidad en múltiples dominios. La creciente complejidad de estas redes ha dado lugar al desarrollo de Deep learning, donde miles de neuronas artificiales distribuidas en múltiples capas son entrenadas para resolver problemas de alta complejidad.

Figura 5: Modelo de la neurona artificial.



Modelo de neurona artificial

Fuente: Elaboración propia.

En la figura que aparece a continuación se muestran las cuatro funciones fundamentales que están relacionadas con la conformación de los modelos

actuales que explican y determinan la denominada Inteligencia Artificial. A continuación, pasamos a describir cada una de ellas, de manera muy breve, tomando en cuenta las respectivas fuentes o referencias bibliográficas consultadas al respecto.

4. Las cuatro funciones más importantes para la Inteligencia Artificial

a) La función lineal

La función lineal es una función de activación que responde proporcionalmente a la entrada, expresada como $f(x)=x$ (ver figura 6a). Su comportamiento no introduce linealidades en la red, lo que limita la capacidad de aprendizaje de patrones complejos. Aunque puede ser útil en la capa de salida de redes diseñadas para problemas de regresión, no permite a las redes profundas modelar relaciones no lineales, lo que la hace poco efectiva en tareas que requieren representación jerárquica de datos (Haykin, 2009).

b) La función de umbral

La función de umbral es una de las funciones de activación más básicas y fue utilizada en los primeros modelos de redes neuronales, como el perceptrón. Su comportamiento binario permite activar una neurona sólo si la entrada supera cierto umbral predeterminado, devolviendo un valor de 0 o 1 (ver figura 6b). Aunque es conceptualmente sencilla y computacionalmente eficiente, su falta de diferenciabilidad impide el uso del algoritmo de retropropagación, lo que limita su aplicabilidad en redes profundas, tal como lo explica Rosenblatt (1958).

c) La función ReLu

La función ReLU (Rectified Linear Unit) es actualmente una de las más utilizadas en redes neuronales profundas debido a su sencillez y eficiencia. Se define como $f(x) = \max(0,x)$, lo que implica que solo las entradas positivas pasan sin modificación, mientras que las negativas se transforman en cero. Esta función acelera la convergencia del modelo y ayuda a mitigar el problema del gradiente desvanecido.

Sin embargo, presenta el inconveniente de que algunas neuronas pueden dejar de activarse durante el entrenamiento, fenómeno conocido como „neurona muerta“, tal como lo muestran ampliamente Goodfellow, Bengio y Courville (2016). En la figura 6c se muestra la representación gráfica de la función ReLu.

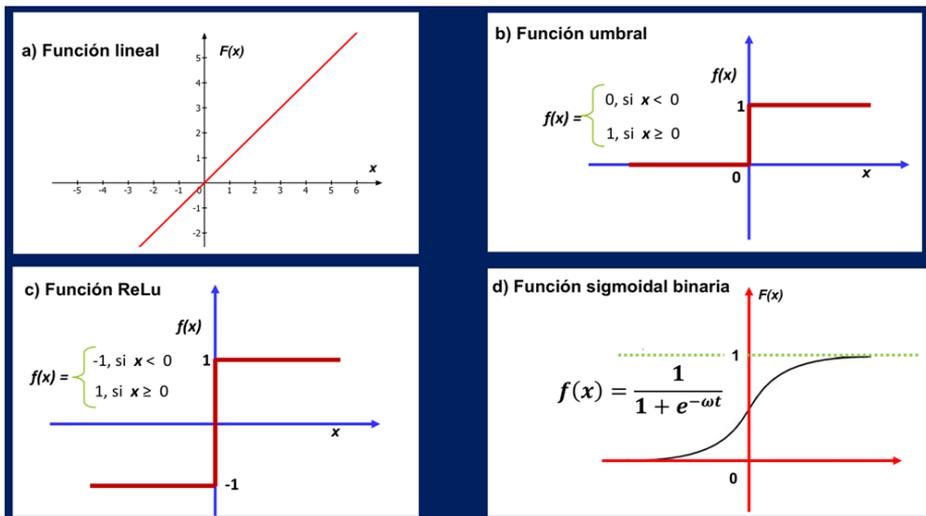
d) La función sigmoidea

La función sigmoidea, también conocida como función logística, transforma la entrada en un rango entre 0 y 1 utilizando la fórmula $f(x) = 1/(1 + e^{-wx})$.

Esta propiedad la hace útil para problemas de clasificación binaria. Su derivabilidad permite aplicar algoritmos de optimización como el descenso del gradiente, aunque sufre del problema del *vanishing gradient* (gradiente decreciente) en redes profundas, lo cual puede dificultar el entrenamiento eficaz (Bishop. 2006).

La función sigmoidea se representa gráficamente en la figura 6d.

Figura 6: Las cuatro funciones matemáticas básicas de la IA



Fuente: Elaboración propia.

Relación entre estas cuatro funciones

a) Introducen no linealidad en los modelos de IA

En la Inteligencia Artificial, especialmente en el aprendizaje profundo (deep learning), las redes neuronales están compuestas por capas sucesivas de nodos o neuronas. Si todas las funciones utilizadas fueran lineales, la red completa se comportaría como una transformación lineal, lo cual limita su capacidad para aprender patrones complejos. Las funciones de activación introducen no linealidad, lo que permite a la red aprender relaciones mucho más ricas y resolver problemas de clasificación, predicción, reconocimiento de voz, imágenes, lenguaje, entre otros.

b) Determinan cómo se transmite la información entre capas

Cada función de activación decide cuánto de la señal (o valor de entrada) se transmite a la siguiente capa. Por ejemplo, ReLU filtra las señales negativas y deja pasar solo las positivas, lo que ayuda a reducir el costo computacional y mejora la eficiencia del entrenamiento. La función sigmoidea, en cambio, es útil cuando se requiere una salida que represente una probabilidad. Así, el tipo de función de activación determina en gran medida el comportamiento interno del modelo de IA, su velocidad de aprendizaje, su precisión y su capacidad de generalización.

c) Permiten el aprendizaje automático mediante retropropagación

La mayoría de los modelos de IA modernos se entrenan usando algoritmos de optimización basados en gradientes, como el descenso del gradiente. Para que estos algoritmos funcionen, se requiere que las funciones de activación sean derivables, de modo que se pueda calcular cómo ajustar los pesos de la red para reducir el error. Funciones como la sigmoidea y ReLU son diferenciables (excepto en puntos específicos), lo cual las hace compatibles con la retropropagación del error, que es el corazón del aprendizaje supervisado en redes neuronales.

5. Conclusión preliminar

Las redes neuronales artificiales se inspiran en el funcionamiento del cerebro humano, en particular en la forma en que las neuronas biológicas se comunican entre sí mediante impulsos eléctricos.

No obstante, en lugar de recurrir a procesos biológicos, las redes neuronales artificiales utilizan estructuras matemáticas y algoritmos computacionales para procesar información, aprender patrones y tomar decisiones basadas en datos. Cada red está compuesta por capas de nodos o „neuronas artificiales“ interconectadas, donde cada nodo aplica una función de activación para determinar la intensidad con la que una señal debe transmitirse a la siguiente capa.

A través de un proceso conocido como entrenamiento supervisado, estas redes ajustan internamente sus pesos y parámetros mediante algoritmos de optimización como el descenso del gradiente, buscando minimizar el error entre las predicciones del modelo y los valores reales observados.

Este proceso iterativo les permite aprender de los datos, es decir, generalizar a partir de ejemplos previos para realizar inferencias sobre nuevos casos. La elección adecuada de las funciones de activación —como ReLU, sigmoidea,

lineal o umbral— resulta esencial para garantizar que la red pueda captar relaciones no lineales y complejas en los datos de entrada.

Gracias a esta capacidad de aprendizaje adaptativo, las redes neuronales artificiales se han convertido en el núcleo de numerosos sistemas de inteligencia artificial avanzados.

Actualmente, son ampliamente utilizadas en aplicaciones de alto impacto como el reconocimiento de voz y de imágenes, la predicción del clima, el *diagnóstico médico* asistido por computadora, los sistemas de recomendación, la traducción automática, entre otros. Su eficiencia y versatilidad permiten resolver problemas que antes requerían intervención humana, transformando múltiples campos del conocimiento y la tecnología.

6. Ejemplo de la aplicación de la IA en el campo de la medicina

Un ejemplo muy conocido sobre los logros, históricamente vistos, de la IA en el campo de la medicina, consiste en la aplicación de las redes neuronales en diagnósticos médicos. En el 2017 un algoritmo superó a radiólogos en detectar el cáncer de mama; muchos/as vieron el futuro de la medicina directamente relacionada con la IA. La pregunta que actualmente están tratando de responder los/as matemáticos/as e informáticos/as, pero también con otros/as profesionales, como personal médicos y paramédicos, etc., consiste en comprender cómo aprenden estas redes a ver y analizar en imágenes lo que los seres humanos han aprendido y practicado reiteradamente durante muchos años. Podríamos señalar que el origen de esta tecnología tiene como antecedentes a McCulloch y Pitts cuando en 1943 lograron modelar neuronas artificiales; sin embargo, para ese entonces tales modelos no podían aprender todavía, hasta que Frank Rosenblatt (1958) desarrolló 14 años después, en 1957, el famoso perceptrón, mencionado al inicio del presente documento.

Por supuesto que los/as matemáticos/as e informáticos/as se enfrentaban ante un gran desafío: las primeras redes necesitaban millones de ejemplos etiquetados, algo complicado y exigente, también en el campo de la medicina. La solución vino posteriormente con el *aprendizaje por transferencia*, lo cual consiste en redes preentrenadas en imágenes genéricas, lo cual pueden ser posteriormente ajustadas mediante un conjunto finito de escáneres médicos. Es así como el DeepMind Health de Google logró hace casi cinco años (2020) predecir enfermedades oculares con un 94% de precisión. Pero un estudio en The Lancet logró mostrar que el programa fallaba con pacientes de grupos étnicos diferenciados, lo cual requería una mayor cantidad de datos e informaciones que pudiesen cubrir buena parte de la población mundial.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

En la actualidad, los/as investigadores/as y aplicadores/as de la IA prometen que estamos al umbral de la superación del ser humano en el ámbito de la predicción y diagnósticos de muchas enfermedades, particularmente en la detección de enfermedades como el cáncer de mama, de piel, etc. Mientras tanto, la Organización Mundial de la Salud, advierte sobre la potencialidad de ciertos sesgos o equivocaciones algorítmicas, especialmente cuando nos atenemos a los diagnósticos automáticos o no humanos. La pregunta central que se hacen quienes tienen tales dudas es la siguiente: ¿Debemos regular tales sistemas como dispositivos médicos para el diagnóstico o verificación de la existencia de enfermedades?

Hay quienes consideran, obviamente, que estas redes y programas neuronales complejos superarán a los médicos, mientras que otros/as piensan que ello no será posible, pero sí tendrá lugar, como ha ocurrido históricamente en la ampliación de la tecnología en la medicina, un reinvento de la praxis medicinal. Una de estas opiniones es la de Fei-Fei Li, quien considera que la IA en medicina debe aumentar la precisión, eficiencia y eficacia del diagnóstico y tratamiento médico, no automatizar o sustituir al ser humano en su actividad cotidiana del ejercicio de la medicina.

Para finalizar este ejemplo, es muy importante resaltar que la IA consiste, por excelencia, en un campo de la investigación y aplicación tecnológica altamente interdisciplinario, donde se requiere, además de la IA, la participación de muchos profesionales humanos, quienes pueden suministrar empatía y respuesta no mecánicas ni automáticas a los pacientes, quienes en esencia son también seres humanos. A pesar de estos importantes adelantos y promesas, hay muchas voces, en el ámbito mundial, que nos advierten sobre los peligros, las falacias, las contradicciones e inconsistencias científicas y filosóficas de la llamada inteligencia artificial, tal como lo señala Lenzen (2020, 9):

„Los críticos, por su parte, temen que la nueva tecnología nos dé sobre todo más oportunidades de vigilancia y manipulación, aumente los monopolios existentes, incremente el desempleo, contribuya a que los procedimientos democráticos se vean cada vez más socavados, provoque el fin del principio de solidaridad en las compañías de seguros y ahonde la división social. Tal vez la tecnología inteligente pueda incluso superarnos algún día, independizarse y, en el peor de los casos, volverse contra nosotros. De hecho, la inteligencia artificial tiene el potencial de confundirnos más que casi cualquier otra tecnología: inevitablemente miramos los algoritmos, los chatbots y los robots a través de la lente de toda la ciencia ficción que hemos visto o leído. Y estas historias están llenas de autómatas con apariencia humana que superan a su creador y se vuelven contra él, y de las irritaciones que pueden surgir cuando humanos y máquinas con apariencia humana se encuentran. Hay máquinas que se confunden con humanos (y viceversa), máquinas que

de repente adquieren conciencia, personas que se enamoran del robot que compraron como esclavo doméstico, etc“.

7. Algunas posiciones a favor de la utilización de la IA en todos los ámbitos del mundo de vida social, productiva, cultural científico y educativo

Sin duda que la llamada inteligencia artificial (IA) ha dejado de ser una tecnología del futuro para convertirse en una realidad cotidiana que transforma múltiples aspectos de nuestra vida diaria. Desde funciones básicas como la asistencia virtual en nuestros teléfonos móviles, hasta procesos complejos en la medicina o la industria, los sistemas de IA se posicionan como herramientas fundamentales para la eficiencia, accesibilidad y desarrollo social. A continuación, quisiera mostrar sólo un conjunto de aspectos positivos que señalan muchos/a autores/as que han venido trabajando con la idea de la aplicación masiva de la IA en todos los campos del mundo de vida y producción de los seres humanos.

En primer lugar, nos encontramos con la automatización y asistencia en tareas cotidianas y complejas. Uno de los aportes más evidentes de la IA es su capacidad de automatizar tareas repetitivas y resolver problemas que probablemente tardarían mucho tiempo en que las personas los solucionen apropiadamente y con alta precisión. Sistemas de aprendizaje automático permiten que algoritmos adapten sus respuestas al contexto y mejoren con el tiempo. Esto resulta crucial no solo para actividades simples (como clasificar correos electrónicos o recomendar música), sino también para áreas especializadas como la detección temprana de enfermedades (Topol, 2019) o la gestión financiera avanzada (Agrawal, Gans y Goldfarb, 2018).

En segundo lugar, se considera que la accesibilidad masiva de la población a datos e informaciones, pero también múltiples beneficios tecnológicos, constituye una de las grandes ventajas de la AI. La popularización de los teléfonos inteligentes, el acceso a internet y la integración de asistentes virtuales como Siri, Alexa o Google Assistant (en el mundo occidental) ha permitido que amplios sectores de la población accedan a las funcionalidades de la IA sin necesidad de conocimientos técnicos avanzados. Este fenómeno democratiza el uso de la tecnología y la convierte en un instrumento de inclusión digital (Brynjolfsson y McAfee, 2014).

En tercer lugar, las traducciones, los procesamientos de voz, etc. permiten el tratamiento de grandes cantidades de datos e informaciones. La velocidad de procesamiento es uno de los pilares del desarrollo de la IA. Al interactuar con alguna de estas ofertas, no siempre gratis, podemos constatar cómo las mismas responden, a su manera obviamente, nos responde sin titubeo a interrogantes sencillas, medianamente difíciles o, inclusive, complejas.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

Algoritmos de análisis de datos pueden procesar millones de entradas en segundos, una tarea imposible para los seres humanos, obviamente. Esto se traduce en eficiencia operativa, reducción de costos y una mejora sustancial en la toma de decisiones, especialmente en sectores como la logística, la salud o el comercio electrónico (Domingos, 2015).

En cuarto lugar, se señala que la IA podría mejorar considerablemente la movilidad y sostenibilidad urbana. Ella se aplica, inclusive en la actualidad, también en el diseño de ciudades más inteligentes. Por ejemplo, la optimización del tráfico mediante análisis predictivo, el uso de sensores y la gestión en tiempo real de rutas puede reducir accidentes, congestión y emisiones contaminantes (European Commission, 2020). Estas tecnologías también se extienden a la logística y las cadenas de suministro, ayudando a minimizar tiempos de entrega y el consumo de energía.

En quinto lugar, pareciera que la IA también contribuye definitivamente en la reducción de la jornada laboral y el equilibrio personal. El economista Christopher Pissarides, Nobel de Economía, ha argumentado que el uso extensivo de la IA en tareas repetitivas podría permitir una semana laboral de cuatro días, sin pérdida de productividad. Esto ofrecería a las personas mayor tiempo libre y posibilidades de desarrollo personal, contribuyendo a un mejor equilibrio entre la vida laboral y familiar (Pissarides, 2023).

En sexto lugar, pareciera que la investigación criminal y análisis de seguridad se beneficiaría altamente con la aplicación adecuada y masiva de la IA, puesto que los algoritmos que la sustentan son cada vez más utilizados para identificar patrones en grandes volúmenes de datos dentro del ámbito de la seguridad. Esto permite detectar conductas sospechosas y prevenir delitos con más rapidez. Sin embargo, este uso debe ser objeto de una evaluación ética rigurosa, pues los sistemas de IA pueden replicar sesgos racistas o clasistas si aprenden a partir de datos contaminados o discriminatorios (Eubanks, 2018; Noble, 2018). Aquí es importante señalar, de manera inmediata, de acuerdo con la literatura crítica hacia la IA, que existe el riesgo de que ésta reproduzca patrones de criminalización selectiva. Por ejemplo, estudios han demostrado que ciertos grupos sociales están sobre-representados en las bases de datos policiales debido a prácticas de vigilancia racista o clasismo institucional. La IA, al aprender de estos datos, puede fortalecer estas tendencias, lo que lleva a una falsa „objetividad algorítmica“ (O’Neil, 2016), la cual obviamente nunca habrá, puesto que ellos carecen de criterios críticos y éticos propios.

En séptimo lugar, es importante destacar el papel emergente y prometedor de la inteligencia artificial (IA) en el ámbito de la agricultura, tanto en la producción como en el procesamiento y distribución de bienes y recursos. Mediante el uso de sensores inteligentes, drones agrícolas, sistemas de

predicción climática y algoritmos de análisis de datos, la IA permite optimizar el uso del agua, fertilizantes y pesticidas, mejorar el rendimiento de los cultivos y reducir las pérdidas de importantes volúmenes de cosechas. Esta capacidad transformadora no se limita a la agricultura, sino que se extiende a múltiples sectores clave como la salud, la educación, la energía, el transporte, la ciencia y la gestión pública. La inteligencia artificial está reconfigurando no solo los procesos técnicos, sino también las estructuras sociales y económicas, generando nuevas dinámicas de innovación, adaptación y transformación en las realidades tanto naturales como sociales (Schmidt y Cohen, 2013; Russell y Norvig, 2021). En este sentido, se insiste en que su potencial no radica únicamente en la automatización o simplificación de tareas, sino en su capacidad para abrir nuevas posibilidades de desarrollo sostenible, equidad social y progreso científico. No obstante, para que este avance tecnológico resulte verdaderamente beneficioso, es imprescindible acompañarlo de marcos normativos sólidos y de políticas éticas, educativas y regulatorias que garanticen un uso inclusivo, justo y responsable de la IA. Sin una supervisión crítica y una planificación cuidadosa, existe el riesgo de que la IA refuerce desigualdades existentes o introduzca nuevas formas de exclusión digital (Floridi et al., 2018).

Por último, no quisiera dejar pasar por alto el surgimiento de la controversia entre la idea del joven matemático Alan Turing y el para entonces también joven filósofo, John Searle, como un argumento expresado por muchos/as defensores/as de la IA, a favor tanto del desarrollo masivo y la aplicación ciega en todos los ámbitos de la vida social, humana y natural. Es decir, el argumento de la *Habitación China*, propuesto por John Searle (1980), plantea un desafío conceptual al famoso test de Turing. Searle argumenta que, aunque una máquina pueda superar el test de Turing —es decir, mantener una conversación indistinguible de la humana— esto no implica que realmente *comprenda* el lenguaje. Según él, la comprensión implica conciencia e intencionalidad, cualidades que no pueden surgir únicamente de la manipulación de símbolos sintácticos. Para Searle, los programas informáticos carecen de semántica: operan con reglas formales, pero sin comprensión del contenido (Searle, 1980).

En contraposición, los defensores del test de Turing (Turing, 1950; Dennett, 1991) sostienen que la inteligencia no debe limitarse a procesos internos de conciencia o subjetividad, sino evaluarse por el comportamiento observable. Desde esta perspectiva funcionalista, si una máquina actúa de manera inteligente, entonces debe considerarse *inteligente* en términos operativos, independientemente de su carencia de conciencia.

Este debate ha cobrado nueva relevancia con el surgimiento de los modelos generativos de lenguaje amplio, conocidos como „GoLLeM“ (por sus siglas en inglés, *Generative Large Language Models*). Estos modelos, como GPT-4,

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

aprenden patrones y asociaciones a partir de enormes corpus de texto, lo que les permite generar respuestas sorprendentemente coherentes y contextuales. A diferencia del „Cuarto Chino“, donde las reglas son explícitas, los GoLLeM operan sobre parámetros probabilísticos que no han sido programados directamente para cada escenario específico. Esto ha llevado a algunos investigadores a sugerir que estos modelos presentan una forma emergente de *protointeligencia* (Bengio, 2023; Bubeck et al., 2023). Sin embargo, sigue siendo debatible si esta capacidad generativa equivale a comprensión. Los GoLLeM carecen de *experiencia subjetiva* y *autoconciencia*, elementos que muchos filósofos asocian con la conciencia plena. Por ello, sus respuestas, aunque sofisticadas, podrían seguir siendo el producto de correlaciones estadísticas, más cercanas al funcionamiento del „Cuarto Chino“ que a un pensamiento verdaderamente humano.

A pesar de estas limitaciones, la inteligencia artificial está revolucionando nuestra forma de trabajar, pensar y crear. Puede procesar cantidades masivas de datos, detectar patrones complejos y automatizar tareas repetitivas, liberando así a los seres humanos para enfocarse en actividades que requieren empatía, juicio ético y creatividad.

No obstante, es fundamental recordar que estos sistemas reflejan los datos con los que fueron entrenados, y por tanto, pueden replicar sesgos sociales y culturales presentes en esos datos (Crawford, 2021). Su uso responsable requiere supervisión humana crítica, marcos éticos sólidos y políticas públicas informadas. Como herramienta, la IA puede ser un catalizador poderoso para la innovación y el bienestar, siempre y cuando se entienda como una *extensión* de la capacidad humana, no como un sustituto de la misma. Usarla para generar ideas, mejorar procesos o tomar decisiones informadas puede potenciar nuestras capacidades, pero jamás debería reemplazar el juicio humano.

8. Algunas posiciones en contra de la aplicación masiva de la IA en todos los ámbitos del mundo de vida social, productiva, cultural científico y educativo

Preliminares: Advertencias sobre la inteligencia artificial y la necesidad de regulación

Algunos expertos en IA llevan tiempo advirtiendo de que el rápido desarrollo de la inteligencia artificial podría dar lugar a programas informáticos autónomos que escapen al control humano (Bostrom, 2014). Se suele decir que las consecuencias podrían ir desde la difusión de información errónea y la pérdida de empleos a gran escala, tal como lo señalan, por ejemplo, Frey y Osborne (2017), hasta escenarios más extremos, como la amenaza a la

seguridad humana (Russell, 2019). Esta es una de las razones por las que los gobiernos están trabajando en el establecimiento de normas para el desarrollo de los diversos softwares que sustentan tecnológicamente a la IA (European Commission, 2021a y 2021b). OpenAI se considera pionera en este ámbito con el software que hay detrás de ChatGPT, aunque su enfoque ha generado tanto curiosidad como preocupación (Metz, 2023).

Algunos expertos sospechan incluso que detrás de estas exhortaciones hay una maniobra deliberada de distracción: los creadores de las IA estarían llamando la atención sobre un hipotético escenario de terror para alejarse de su responsabilidad en los peligros más inmediatos, como las noticias falsas automatizadas (Vosoughi et al., 2018), la pérdida de empleos debido a la automatización (Acemoglu y Restrepo, 2020) y la erosión de la confianza en las instituciones (Tucker et al., 2018). Todo ello es grave, pero probablemente no tan devastador como los escenarios apocalípticos que a veces se esbozan, sin importar mayores análisis académicos y científicos (Brooks, 2023).

Empero, existe un argumento a favor de estas consideraciones alarmantes: la propia naturaleza impredecible de los sistemas de la IA. Cuando los/as desarrolladores/as entrenaron a sus modelos para entender el lenguaje humano, estos se volvieron tan complejos que incluso sus propios/as creadores/as no pueden predecir con certeza y convicción sus capacidades (Marcus, 2022).

Esto ha llevado a un enfoque basado en los viejos procedimientos experimentales del ensayo y error, donde la incertidumbre es la que juega el papel central de la experimentación (Amodei et al., 2016). Por ello, las exhortaciones pueden parecer vagas, pero reflejan la dificultad de evaluar riesgos en sistemas cuyos comportamientos emergentes no son del todo comprensibles (Bender et al., 2021). Sin embargo, si las empresas de la IA realmente tomaran en serio su llamado a la precaución, deberían dar el primer paso y abrir sus modelos a un escrutinio masivo, participativo e independiente, tal como lo exigen, entre otros/as, Brundage et al., 2020. Hasta ahora, esto no ha ocurrido.

Un ejemplo claro es Sam Altman, es el director ejecutivo (Chief Executive Officer) de OpenAI, la empresa de inteligencia artificial detrás de modelos como ChatGPT, GPT-4 y DALL-E. Sus declaraciones y decisiones influyen en cómo se desarrolla y regula la IA a nivel global, siendo uno de los firmantes de cartas pidiendo regulación, cuya empresa mantiene en secreto detalles clave de GPT-4, como su arquitectura, hardware y datos de entrenamiento, tal como lo indicó recientemente Hao (2023). OpenAI justifica esta opacidad por motivos de seguridad y competitividad, pero ello contradice su retórica sobre transparencia y responsabilidad (Stokel-Walker, 2023).

9. Los Inconvenientes de la Inteligencia Artificial: Un Análisis Crítico y Reflexivo

I. Falta de empatía y conciencia social

Una de las principales desventajas de la inteligencia artificial (IA) es la falta de empatía y conciencia social, elementos esenciales para la toma de decisiones éticas y morales en contextos humanos. La IA no posee emociones ni capacidades cognitivas que le permitan interpretar la complejidad de las relaciones humanas. Por esta razón, no puede realizar juicios morales auténticos, sino que reproduce las estructuras de poder y las normas que existen en los datos con los que es entrenada. Las máquinas, en lugar de integrar consideraciones éticas, se limitan a optimizar funciones basadas en algoritmos y parámetros preestablecidos.

Este comportamiento puede resultar en decisiones que, aunque técnicamente correctas, carecen de la sensibilidad humana necesaria, especialmente en situaciones que involucran dilemas éticos complejos, tal como lo analiza muy acertadamente Binns (2018). Si bien sus argumentos teóricos tienen casi siete años de haber sido publicados y la IA ha avanzado mucho durante los últimos cinco años, se podría seguir considerando que la falta de empatía y conciencia social que la caracteriza será un gran dilema en el transcurso de las próximas décadas.

Por otra parte, debido a que la IA no tiene un sentido de responsabilidad social o moral, tiende a perpetuar las desigualdades y las relaciones de poder que existen en los datos de entrada y la propia producción textual generada por ella. Este problema se evidencia en el uso de la IA en sistemas judiciales, de contratación laboral, e incluso en la política y funcionamiento de los Estados-Naciones, donde los algoritmos pueden contribuir a sesgos de discriminación, sin ser conscientes de sus implicaciones sociales, tal como lo explica en su amplio trabajo Zuboff (2019). La incapacidad de las máquinas para interpretar el contexto humano y social, particularmente desde el punto de vista de posicionamientos y criterios críticos, tal como lo hacemos los seres humanos, constituye un desafío significativo, especialmente en decisiones que deben ser equitativas y justas para todos los involucrados.

II. Impacto en el Mercado Laboral: Transformación Disruptiva

El impacto de la IA en el mercado laboral es otro de los grandes inconvenientes que genera preocupación. A medida que la tecnología avanza, ciertos trabajos tradicionales están siendo reemplazados o modificados por sistemas automáticos que pueden realizar tareas previamente desempeñadas por los humanos. Profesiones como el periodismo, la educación, la medicina y el derecho enfrentan una transformación radical debido a la automatización de

procesos, los cuales estas máquinas pueden hacer rápida y masivamente. En particular, los sistemas de IA como ChatGPT, DeepSeek, entre otros, pueden generar contenidos automáticamente, lo que ha llevado a que algunas áreas del periodismo, como la curaduría de noticias (la cual hace referencia a la selección, filtrado y presentación de noticias por parte de editores), haya quedado relegada por algoritmos o plataformas que destacan la información relevante, veraz o de interés para una audiencia específica.

Estas y muchas otras profesiones experimentarán una disminución de la demanda laboral de las mismas, tal como lo muestran Brynjolfsson y McAfee (2014). Aunque actualmente tales sistemas complejos de IA no tienen la capacidad de reemplazar por completo el trabajo humano en campos como el periodismo, la construcción, la agricultura, etc., el ritmo de avance tecnológico indica que en el futuro las máquinas podrán realizar tareas más complejas que hoy están reservadas exclusivamente a la acción y el pensamiento de los seres humanos.

A pesar de que la automatización, según sus defensores/as, podría crear nuevos tipos de empleos, también plantea la desaparición de muchos otros, particularmente aquellos que requieren de un alto grado de creatividad o empatía humana. Este cambio disruptivo y destructivo también puede generar una brecha significativa entre aquellos que tienen acceso a las nuevas tecnologías y aquellos que no, exacerbando la desigualdad social y económica entre las personas, los países y las regiones (Frey y Osborne, 2017). La reconfiguración del mercado laboral será inevitable y, aunque se podrían generar nuevas oportunidades, estas no necesariamente reemplazarán los puestos de trabajo eliminados, lo que añade complejidad a la cuestión de la equidad y la distribución de la riqueza, produciéndose claramente una mayor diferencia entre quienes poseen las riquezas y las grandes mayorías que viven en la pobreza, tal como lo vemos claramente en la actualidad.

III. La Veracidad de la Información: Un Desafío Epistemológico

Un problema fundamental con los sistemas de IA, especialmente aquellos basados en modelos generativos como ChatGPT, DeepSeek, entre otros, es que no están diseñados para verificar la veracidad de la información. Estos modelos operan mediante algoritmos estadísticos de predicción lingüística, lo que significa que generan respuestas basadas en patrones y probabilidades extraídas de grandes volúmenes de datos e informaciones. Sin embargo, estos datos e informaciones no siempre provienen de fuentes confiables o verificadas.

Por lo tanto, los sistemas de IA pueden difundir información incorrecta o sesgada sin tener la capacidad de evaluar su exactitud, tal como lo han demostrado Bender et al., (2021). Este problema se ve agravado por el hecho

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

de que muchos de los datos e informaciones utilizadas provienen de Internet, donde abundan tanto fuentes legítimas como desinformación. Uno de los grandes problemas que tienen tales sistemas tiene que ver con el escaso o poco acceso a fuentes referencias bibliográficas que han sido publicadas en formato físico, sin existencia del formato digital, o que sencillamente no han tenido la posibilidad de ser digitalizados por cuestiones de carácter académico, científico, económico e intelectual.

La dificultad para discernir entre hechos verídicos y falsos en la era digital se ve intensificada por el rol de la IA en la propagación de contenidos, muchos brillantes, unos ética y científicamente cuestionables y una gran mayoría altamente problemáticos. A medida que la IA se convierte en una fuente importante de información, los usuarios pueden asumir erróneamente que las respuestas generadas por estas máquinas son infalibles. La confianza en los sistemas actuales de IA como generadores de contenido autoritativo puede tener consecuencias graves, especialmente cuando se trata de temas delicados como la política, la salud, la educación-formación o la ciencia, tal como lo indica O'Neil (2016). Este fenómeno de „epistemología de la IA“ nos recuerda que la veracidad no puede depender exclusivamente de la tecnología, sino que debe ser un esfuerzo humano de verificación y reflexión crítica.

IV. La Amenaza de la Falsificación: Falsificaciones y Manipulación de la Realidad

Otro de los riesgos más alarmantes de la IA es su capacidad para crear contenido falso de alta calidad, como imágenes, videos y audios, que pueden ser indistinguibles de los originales. Este fenómeno, conocido como „deepfakes“ (falsificaciones profundas), ha demostrado ser una herramienta poderosa para la creación de contenidos manipulativos y altamente discutibles. Cuando este tipo de contenido es distribuido sin ser identificado como falso, puede tener un impacto destructivo, manipulativo, desorientador y controlador en el tejido social y político, especialmente en momentos de alta tensión, como durante elecciones o crisis políticas (Mora, 2020; Chesney y Citron, 2019a y 2019b). La manipulación de la opinión pública a través de falsificaciones y manipulación de la realidad puede socavar los principios fundamentales del debate político, democrático y sociocultural, ya que los/as ciudadanos/as pueden ser engañados/as por una realidad distorsionada que afecta la toma de decisiones de manera independiente, crítica y participativa.

La propagación de deepfakes plantea un desafío de regulación y control, ya que la velocidad a la que los contenidos falsos pueden viralizarse es impresionante, desgastante y aberrante, tanto de la conciencia social como de la formación integral de buena parte de nuestra población. Además, la capacidad de la IA para generar estos contenidos con un alto nivel de

sofisticación, sin importar su autenticidad y sus graves consecuencias, complica la tarea de identificación y verificación, dejando a los usuarios vulnerables a la manipulación y desinformación, tal como o analizan críticamente Woolley y Howard (2018). En este sentido, la IA no solo está transformando masivamente la producción de contenidos, sino también la naturaleza misma de la realidad, creando un entorno ambivalente en el que la distinción entre lo verdadero y lo falso se vuelve cada vez más difusa.

La Paradoja de la Tecnología y sus alcances imprevisibles

En definitiva, la inteligencia artificial presenta una paradoja en nuestra era tecnológica. Por un lado, la IA ofrece soluciones eficientes a muchos problemas cotidianos, desde la compra de boletos hasta la optimización del transporte. Sin embargo, también plantea serios desafíos, especialmente en términos de la autonomía humana, la toma de decisiones y la preservación de valores sociales fundamentales. Mientras que algunos argumentan que la tecnología puede reducir los errores humanos y optimizar los procesos, otros advierten sobre la pérdida de creatividad y la individualidad humana, elementos que se ven reemplazados por la eficiencia de las máquinas, tal como lo indica analíticamente Turkle (2011).

Además, como se observa en las teorías de Karl Marx (1867), Max Weber (1905) y David Harvey (2005), la IA no solo afecta la productividad, sino que también se ha convertido en un instrumento para la concentración de poder, especialmente en el capitalismo basado y orientado en tales plataformas informáticas. Las grandes corporaciones tecnológicas como Google, Amazon, Facebook (Meta), Apple, Microsoft, etc., están utilizando enormes cantidades de datos e informaciones, lo que les otorga un poder desmesurado que podría ser difícil de regular, sin la intervención activa y decisiva de los gobiernos respectivos, en el ámbito global, tal como lo indican Zuboff (2019), Mora (2020) y Wind y Mora (2023a y 2023b), y la acumulación de datos por parte de estas empresas genera dinámicas de monopolización y exclusión que refuerzan las desigualdades sociales y económicas (Srnicek, 2017).

En última instancia, la inteligencia artificial, aunque poderosa, no puede resolver los problemas fundamentales de nuestro tiempo, como la desigualdad socioeconómica de gran parte de la población o el cambio climático. Lo que puede hacer es acelerar procesos, aumentar la eficiencia y modificar las relaciones de poder, pero siempre dentro de un contexto que debe ser vigilado cuidadosamente por los humanos que diseñan, implementan y utilizan estas tecnologías (Mora, 2020; Wind y Mora, 2023a).

Es importante aclarar que la ausencia de creencias morales en la IA no implica que sea moralmente indiferente. Aunque los sistemas de la IA están diseñados para realizar tareas específicas y proporcionar información, o asistir

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

en la toma de decisiones basadas en los datos con los que han sido entrenados, la responsabilidad ética recae en los humanos que los desarrollan y utilizan. Ellos deben asegurarse de que dichos sistemas operen de acuerdo con principios morales y éticos, tal como lo indican con claridad y precisión Binns (2018) y O'Neil (2016).

La moralidad de la terraformación de Marte (proceso de transformar un planeta o satélite natural para que sea habitable para los seres humanos, creando condiciones similares a las de la Tierra) es un tema profundamente complejo y debatido entre científicos, filósofos, sociólogos, psicólogos, juristas, entre muchos/as otros/as profesionales.

Algunos/as argumentan que sería moralmente justificable transformar Marte para crear un entorno habitable para los seres humanos y otras formas de vida, mientras que otros/as sostienen que es moralmente incorrecto alterar el estado natural de un planeta. Por el contrario, es más importante y significativo intentar resolver definitivamente los grandes problemas que afectan actualmente a nuestro planeta Tierra, para lo cual la tecnología podría contribuir considerablemente (Mora, 2020 y 2023). En última instancia, la valoración moral de la terraformación dependerá de la perspectiva de cada individuo sobre la importancia de preservar el estado natural de los planetas y las posibles consecuencias de intervenir en su ecosistema (Borenstein et al., 2017; Hall, 2013).

Un argumento en favor de preservar el estado natural de otros planetas es que la diversidad e integridad de los ecosistemas cósmicos son valiosas. Estos han existido durante millones de años y poseen características y propiedades únicas que podrían ser destruidas o alteradas irreversiblemente a través de la intervención humana (Zuboff, 2019; McKibben, 2017), lo cual realmente no es necesario ni indispensable, cuando uno observa que las personas que muy probablemente se beneficiarían de tales complejas acciones serían sólo una minoría, aquellos que disponen de gran poder y grandes recursos económicos.

Además, se argumenta que la terraformación de otros planetas podría generar consecuencias imprevistas, como la alteración del clima o la atmósfera de una manera que resulte peligrosa o perjudicial para las formas de vida potenciales que pudieran existir en ellos (Chesney y Citron, 2019a y 2019b). Sin embargo, también hay quienes defienden que los beneficios potenciales de terraformar otros planetas, como la creación de nuevos entornos habitables, superan los riesgos, y que la humanidad tiene una responsabilidad moral de explorar y utilizar los recursos planetarios para su propio beneficio, tal como lo indica, por ejemplo, Bostrom (2014).

Como sistema complejo, físico y mecánico, la IA no posee una perspectiva personal. No obstante, la misma puede proporcionar información y análisis sobre un tema determinado, lo cual hace obviamente muy bien. Como modelo de aprendizaje automático, ha sido entrenado con grandes volúmenes de datos numéricos e informaciones textuales, pero carece de experiencias, sentimientos y conciencia personal. No es autoconsciente ni capaz de desarrollar perspectivas propias. Su capacidad se limita a proporcionar información y análisis (muy rápido y eficiente, por supuesto) en función de los datos con los que ha sido entrenado, sin la habilidad de formular opiniones o creencias personales. Aunque entiende e interpreta el lenguaje, no posee la experiencia subjetiva que caracteriza a los humanos (Turkle, 2011). La falta de criterio propio, conciencia crítica, autodeterminación, autonomía mental y conceptual, hace muy difícil creer, pensar y esperar que ella pueda igualar o sustituir a los seres humanos.

La moralidad de recurrir a una IA amoral para la toma de decisiones morales es un tema polémico y complejo, ampliamente debatido por filósofos/as y expertos/as en ética. Algunos/as defienden que es moralmente justificable utilizar la IA en la toma de decisiones éticas, ya que puede ofrecer análisis y conocimientos valiosos que ayuden a los humanos a tomar decisiones más informadas. Además, la IA puede identificar patrones y tendencias que son difíciles de discernir para los seres humanos, y puede colaborar en la minimización de sesgos inherentes a la toma de decisiones humanas, tal como lo explican Brynjolfsson y McAfee (2014).

A manera de concluir este apartado sobre los riesgos y consecuencias negativas de la IA, desarrollando continuación siete aspectos fundamentales:

A. Deepfakes: La manipulación sintética de rostros y voces

Uno de los riesgos más notorios de la inteligencia artificial es el desarrollo y difusión de los llamados deepfakes, contenidos audiovisuales generados por IA que simulan de manera extremadamente realista la apariencia y la voz de personas reales. Esta tecnología representa una amenaza directa a la reputación individual, ya que puede utilizarse para crear videos o audios falsos que difamen o comprometan la imagen pública de cualquier persona.

Además, estas manipulaciones pueden utilizarse para engañar a usuarios mediante videos o llamadas fraudulentas, lo que ha dado lugar a nuevos tipos de extorsión y estafas, muchas a gran escala, incluyendo a una variada y heterogénea población. En los ámbitos político y periodístico, los deepfakes podrían erosionar la confianza en los medios de comunicación, distorsionar la percepción pública y manipular la opinión ciudadana a través de campañas de desinformación sofisticadamente elaboradas, lo cual vemos cada vez con mucha frecuencia y consecuencias desastrosas para los países, las

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

comunidades, las poblaciones, las personas colectiva e individualmente consideradas.

B. Estafas Automatizadas a gran escala

A diferencia de los deepfakes, las estafas automatizadas no siempre requieren material visual o auditivo manipulado, sino que se basan en algoritmos capaces de simular interacciones humanas para engañar a las víctimas. La IA permite automatizar estas operaciones de forma masiva, facilitando fraudes como los conocidos lovescams (estafas románticas), donde los ciberdelincuentes simulan relaciones afectivas para obtener beneficios económicos. Este tipo de prácticas puede escalar sin necesidad de supervisión humana directa, lo que incrementa tanto la frecuencia como la sofisticación de los ataques.

Como consecuencia, tanto individuos como organizaciones enfrentan un aumento significativo del riesgo de pérdidas económicas a causa de fraudes sistemáticos impulsados por la inteligencia artificial y, por supuesto, por toda la tecnología relacionada con la algoritmización del mundo de vida socionatural. Ello tiene que ver con el proceso de usar algoritmos para guiar o influir en nuestras decisiones, especialmente en el ámbito de la tecnología y la gestión de datos. Este fenómeno implica el uso de algoritmos para tomar decisiones en diversas áreas, desde el sector empresarial hasta la vida cotidiana, y puede tener implicaciones importantes en la sociedad.

C. Malware Inteligente: Adaptación y resiliencia algorítmica

Otro riesgo crítico lo constituye el malware potenciado por IA, una nueva generación de software malicioso capaz de adaptarse dinámicamente a las medidas de seguridad de los sistemas. Este tipo de malware utiliza algoritmos que analizan el comportamiento del sistema objetivo y modifican automáticamente su código para evitar ser detectado por los antivirus tradicionales. Gracias a herramientas de personalización basadas en IA, como las API de OpenAI, el código puede ajustarse en tiempo real, volviendo obsoletos muchos de los métodos tradicionales de defensa cibernética.

La creciente capacidad de este malware para evolucionar con cada infección convierte su detección y mitigación en un desafío técnico de alto nivel. Se podría decir, que la digitalización tiene en sí misma potencialidad de autodestruirse o por lo menos de hacerse daño así mismo, de manera independiente de la influencia humana o mediante la intervención externa de las personas, especialmente de quienes están dedicados/s, por razones e intereses también económicos y criminales, en impulsar tales destrucciones digitales masivas particulares.

D. Discriminación algorítmica y falta de transparencia

La inteligencia artificial también plantea riesgos importantes en términos de justicia social y derechos humanos, especialmente cuando reproduce sesgos existentes en los datos con los que ha sido entrenada. Este fenómeno, conocido como discriminación algorítmica, puede dar lugar a decisiones injustas en ámbitos críticos como el empleo, la justicia, la salud o la educación. Además, la opacidad en el funcionamiento de muchos sistemas de IA dificulta la rendición de cuentas y socava la confianza pública en su uso.

Las empresas, instituciones, los gobiernos y personas particulares, en general, deben asumir una responsabilidad ética y legal significativa para garantizar que sus sistemas de IA sean transparentes, justos y respetuosos con los derechos fundamentales de todos los seres humanos, ya que un manejo negligente de los datos e informaciones puede dañar no solo las finanzas, la economía, la democracia, las instituciones, la naturaleza, etc., sino también la soberanía, dignidad e independencia humanas. Por supuesto que sería iluso pensar que los/as defensores/as, impulsores/as y fanáticos/as de la IA estén interesados/as en regulaciones y controles estrictos del desarrollo y uso de tales plataformas algorítmicas.

E. La compleja asignación de responsabilidad

La implementación de sistemas autónomos alimentados por IA ha generado un nuevo desafío jurídico: determinar la responsabilidad en caso de fallos o daños. ¿Quién es responsable cuando un vehículo autónomo causa un accidente? ¿El propietario, el fabricante, o el programador? La falta de claridad en este ámbito legal puede reducir significativamente la confianza de los/as usuarios/as en las tecnologías emergentes, las cuales están en boga actualmente.

A su vez, eximir completamente a las empresas de responsabilidad podría fomentar el desarrollo de productos inseguros, mientras que una regulación excesivamente estricta podría sofocar la innovación. En este contexto, es esencial encontrar un equilibrio adecuado que permita garantizar la seguridad y los derechos de las personas, sin desincentivar la investigación y el progreso tecnológico.

A pesar de esta sugerencia algo filantrópica e ingenua, consideramos que la salvación del planeta tierra, la humanidad y las cosas bellas que han caracterizado la vida social y natural, estará en las manos de la población política y críticamente comprometida y no en manos de las grandes empresas, consorcios e industrias productoras de una tecnología contraria a la vida social y natural en este planeta, tal como la inmensa producción automovilística y militar mundial.

F. Nuevos blancos para la ciberdelincuencia

La creciente integración de IA en sistemas digitales ha abierto una nueva puerta de entrada para los ciberdelincuentes. Como señala el experto en ciberseguridad Mirko Hyppönen (2022, 130), „[...] todo lo que se etiqueta como inteligente es vulnerable“, lo que subraya el riesgo que representan estas tecnologías mal protegidas. Uno de los fraudes más preocupantes es el llamado „fraude del CEO“ (Chief Executive Officer), en el cual los atacantes utilizan técnicas de imitación basadas en IA para hacerse pasar por directivos de alto rango y ordenar transferencias fraudulentas.

Este tipo de amenazas no sólo ponen en riesgo información confidencial, sino que también comprometen la estabilidad financiera de las empresas y la confianza en los entornos digitales.

G. Ética y privacidad en la era de la IA

Por último, los riesgos éticos relacionados con el uso de la IA y la protección de datos personales son de una magnitud considerable. La capacidad de los algoritmos para analizar grandes volúmenes de datos puede derivar en usos indebidos, incluso involuntarios, que vulneren la privacidad y los derechos fundamentales de las personas. Asimismo, la falta de transparencia en los procesos de toma de decisiones algorítmicas puede generar desconfianza cuando se trata de asuntos sensibles, como la selección de personal, la concesión de créditos o los diagnósticos médicos.

Estos dilemas éticos exigen marcos regulatorios sólidos y una supervisión constante para garantizar que el desarrollo tecnológico se alinee con principios de equidad, dignidad humana y protección legal. Sobre este tema se ha escrito mucho en diversas lenguas y en muchos países, algunas de las referencias que sustentan los aspectos éticos que debería ser tomando en cuenta en el desarrollo y utilización de la IA, serían los siguientes: Floridi y Cowls, 2019; Eubanks, 2018 y European Commission, 2021a y 2021b).

10. Probables usos de la Inteligencia Artificial Generativa (IAG) en el desarrollo didáctico y formativo

La Inteligencia Artificial Generativa (IAG) está modificando profundamente los escenarios educativos contemporáneos, planteando tanto innovaciones transformadoras como tensiones éticas, pedagógicas y técnicas. Su presencia en el aula ha estimulado debates sobre el papel del docente, la autoría intelectual, la personalización del aprendizaje y la evaluación automatizada.

Según Cabellos, De Aldama y Pozo (2024), muchos/as docentes universitarios se muestran ambivalentes frente al uso de la IAG, reconociendo su utilidad,

pero también expresando preocupación por su impacto en la calidad y el sentido pedagógico de la enseñanza. Esta dualidad exige una mirada crítica y equilibrada que no sólo evalúe su potencial, sino también sus implicaciones estructurales para el aprendizaje.

La educación, la formación, la pedagogía, la didáctica, el aprendizaje y la enseñanza son conceptos profundamente humanos que trascienden el ámbito técnico y requieren una comprensión compleja de las dinámicas sociales, culturales y emocionales involucradas en los procesos educativos y formativos. Estos elementos no sólo se refieren a la transmisión de conocimientos, sino que también están íntimamente ligados al desarrollo integral de las personas, sus capacidades cognitivas, emocionales y sociales.

Como subraya Mora (2017), estos aspectos son demasiado importantes y fundamentales para dejarlos, de manera simple e irresponsable, en manos de la algoritmización del mundo social y natural que predomina en la actualidad. La educación no puede ser reducida a simples procesos automatizados, ya que, de hacerlo, se corre el riesgo de deshumanizar el aprendizaje, ignorando la necesidad de relaciones interpersonales, empatía y contextualización que los/as educadores/as ofrecen en su labor diaria.

A pesar de esta crítica, no se debe interpretar que los/as docentes se opongan al uso de la tecnología digital en el ámbito educativo. Wind y Mora (2023a y 2023b) argumentan que, de hecho, la integración de las herramientas digitales es una parte esencial del contexto educativo contemporáneo. Los avances tecnológicos, como plataformas de aprendizaje en línea, aplicaciones educativas y herramientas de colaboración digital, pueden enriquecer y complementar los métodos tradicionales de enseñanza.

Sin embargo, esta integración debe ser cuidadosamente planificada y gestionada, reconociendo que la tecnología debe ser un medio que potencie, y no que reemplace, la interacción humana y los procesos pedagógicos que implican la reflexión crítica, la adaptación a las necesidades de los estudiantes y el fomento de la creatividad y el pensamiento independiente.

La reflexión crítica sobre el uso de la tecnología en la educación también se extiende al rol de los algoritmos en la personalización del aprendizaje. Aunque la inteligencia artificial y los sistemas adaptativos pueden ofrecer experiencias de aprendizaje más personalizadas, Mora (2017) señala que estos sistemas pueden estar sujetos a sesgos inherentes en los algoritmos que los alimentan, lo que puede generar exclusión o desigualdades en la educación. Así, es fundamental que el diseño y la implementación de la tecnología educativa se realicen con un enfoque ético y consciente de sus implicaciones en la equidad educativa.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

En este sentido, se hace necesario un equilibrio entre la incorporación de herramientas digitales y la preservación de los principios fundamentales de la educación: la equidad, la creatividad, la interacción humana y el desarrollo de habilidades críticas. Wind y Mora (2023a y 2023b) concluyen que, si bien la tecnología tiene un papel importante en la educación moderna, no debe ser vista como un fin en sí misma, sino como un medio para fortalecer los objetivos pedagógicos, siempre bajo la supervisión, el juicio y la intervención activa de los docentes.

Uno de los usos más destacados de la IAG en educación es la personalización del aprendizaje, permitiendo adaptar ritmos, contenidos y métodos a las necesidades individuales de los estudiantes. Esta capacidad, materializada en tutores/as inteligentes y plataformas adaptativas, ha sido valorada positivamente por su capacidad de atender la diversidad educativa. No obstante, autores/as como Floridi et al. (2018) advierten que esta personalización está mediada por algoritmos entrenados con datos e informaciones que pueden contener sesgos, lo que podría reproducir y amplificar desigualdades ya existentes. Además, según Kap (2024), el diseño de estas tecnologías suele carecer de transparencia, dificultando la comprensión del modo en que se toman decisiones sobre el proceso educativo.

En segundo lugar, destaca la creación automatizada de materiales didácticos. Herramientas como ChatGPT, DeepSeek, Midjourney o DALL·E son capaces de generar textos, imágenes y recursos visuales de forma inmediata, facilitando la labor docente y abriendo nuevas posibilidades para el aprendizaje visual y narrativo. Sin embargo, estudios como el de García Martínez y Pujol (2024) subrayan la falta de rigor y precisión en algunos de los contenidos generados, lo cual puede comprometer la calidad educativa. Además, surge el dilema de la autoría: ¿a quién pertenece el conocimiento generado por estas herramientas y qué lugar queda para la creatividad y criterio profesional del docente?

Un tercer ámbito de aplicación es la evaluación automatizada y la retroalimentación inmediata. Los sistemas basados en la IAG pueden calificar tareas, generar rúbricas y ofrecer recomendaciones personalizadas. Esta capacidad acelera procesos y proporciona a los estudiantes respuestas en tiempo real. No obstante, Bond et al. (2024) señalan que estas tecnologías aún presentan limitaciones notables en la valoración de competencias complejas como el pensamiento crítico, la argumentación o la creatividad. Además, existe el riesgo de reducir la interacción humana significativa entre los/as docentes y los/as estudiantes, un aspecto esencial en la formación integral colectiva e individual.

Finalmente, la IAG ha impulsado el desarrollo de asistentes virtuales, chatbots educativos y entornos inmersivos mediante realidad virtual y aumentada. Estas herramientas facilitan el aprendizaje autónomo y permiten realizar simulaciones en entornos seguros, como laboratorios virtuales o prácticas clínicas simuladas.

No obstante, Kap (2024) advierte que estos avances pueden contribuir a la deshumanización de la experiencia educativa si no se integran críticamente, y además pueden presentar barreras de acceso económico para estudiantes y centros educativos con recursos limitados.

Por ello, la implementación de la IAG debe guiarse por principios éticos, justicia educativa y políticas públicas inclusivas que garanticen su uso responsable, equitativo y, sobre todo, transformando las relaciones sociales y económicas de producción y explotación prevalecientes.

11. Del triángulo pedagógico centrado en la IAG al hexágono pedagógico relacionado con la IAG

Aquellos que buscan integrar la Inteligencia Artificial Generativa (IAG) en la enseñanza, utilizando modelos didácticos que han sido desarrollados a lo largo de muchos años por pedagogos, educadores, y diseñadores curriculares, se enfrentan a una realidad compleja. La IAG forma parte de la nueva realidad académica y profesional, especialmente en el campo de la formación inicial de los docentes, así como en las diversas actividades de formación continua dirigidas a maestros, profesores y docentes en general.

De acuerdo con Cabellos, De Aldama y Pozo (2024), la introducción de la IAG en los entornos educativos supone una transformación significativa, pero debe ser abordada con precaución y reflexión pedagógica. Algunos gobiernos, a través de sus respectivos ministerios de educación, han establecido directrices que promueven el uso masivo de sistemas de IAG en las aulas, enfocándose en la medición de rendimientos y proporcionando orientaciones para su implementación a corto y mediano plazo en todos los ámbitos del sistema educativo, sin comprometer la integridad pedagógica, conceptual y científica.

Sin embargo, como señalan Bond et al. (2024), la integración de nuevas tecnologías en el aprendizaje y la enseñanza es un proceso a largo plazo que, en gran parte, se lleva a cabo de forma descentralizada, adaptándose de manera específica a cada asignatura y contexto. En los programas de grado y en los cursos de formación continua para docentes en la mayoría de los países, se insiste en la necesidad de mantener una orientación que guíe la transformación de la educación en la era digital actual, considerando tanto los retos como las oportunidades que la IAG ofrece.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

Por supuesto que la educación está en constante transformación (Mora, 2017). Se dice, por ejemplo, que un mundo cada vez más digitalizado cambia las condiciones marco de la enseñanza, el estudio y el aprendizaje. Las instituciones educativas se enfrentan al reto de cómo ayudar mejor a los/as estudiantes a desarrollar las potencialidades necesarias para prosperar en este entorno.

En este sentido, las habilidades y destrezas digitales y la alfabetización digital se han convertido en componentes fundamentales del currículo educativo (UNESCO, 2021). Los/as educadores/as emplean modelos y conceptos didácticos como herramientas para integrar de forma estructurada los aspectos digitales en las actividades educativas.

Figura 7: el triángulo didáctico de Dagstuhl



Fuente: Döbeli Honegger y Salzmänn (2018, 109).

El Triángulo de Dagstuhl (figura 7), originado en la didáctica de la informática y la educación mediática, fue desarrollado en 2016 en un seminario de GIDagstuhl. Posteriormente, fue revisado por Beat Döbeli Honegger y Renate Salzmänn en 2018 (Döbeli Honegger y Salzmänn, 2018, 109). Este triángulo permite clasificar los fenómenos digitales en tres dimensiones fundamentales: Perspectiva Tecnológica, Perspectiva Sociocultural, Perspectiva Orientada a la Aplicación. Dos críticas fundamentales hacemos a este triángulo pedagógico, una consiste en las tres perspectivas allí establecidas y la segunda tiene que ver con el viejo concepto pedagógico del triángulo didáctico, constituido sólo por docentes, estudiantes y contenidos, el cual hemos analizado y criticado en varias oportunidades (Mora, 2016).

Si bien, se han difundido el triángulo de Dagstuhl (Döbeli Honegger y Salzmann, 2018), como un ejemplo de modelo para categorizar la aplicación de sistemas de IAG, como ChatGPT, DeepSeek, en la enseñanza. A continuación, queremos presentar un resumen sobre dicho concepto didáctico, ampliándolo a un hexágono, con la incorporación de otras tres componentes, la cuales nos parecen más significativas e indispensables para darle sentido sociocrítico, ético-político, biológico neuronal-artificial y productivo transformador de la propuesta de estos/a autores/as. En la figura adjunta se muestra la conjunción de los dos hexágonos que hemos desarrollado para transformar definitivamente la educación, la formación, la pedagogía, la didáctica, el currículo, el aprendizaje y la enseñanza.

El objetivo del Triángulo de Dagstuhl consistió en ofrecer un modelo que ayudará a los/as educadores/as, investigadores/as y diseñadores/as curriculares a reflexionar sobre cómo las tecnologías afectan la educación desde tres grandes perspectivas, promoviendo un enfoque integral y equilibrado en la incorporación de la tecnología en las aulas. El modelo no sólo busca describir las implicaciones de la tecnología en el aprendizaje, sino también fomentar una reflexión crítica sobre cómo la tecnología debe ser utilizada de manera ética y responsable, teniendo en cuenta sus posibles efectos en los individuos y las sociedades.

Este hexágono ofrece un modelo más completo e integrador para analizar la implementación de la IAG en la educación, la formación y la pedagogía en general. La irrupción de la Inteligencia Artificial Generativa (IAG) en el campo de la educación y la formación exige expandir las relaciones tradicionales del triángulo pedagógico hacia una comprensión más compleja y multidimensional, el cual podría estar representado mediante un hexágono pedagógico, tal como se muestra en la figura 8, lo cual será explicado más adelante.

Por supuesto, las propuestas basadas en el famoso triángulo didáctico han sido criticadas en otras oportunidades (Mora, 2010, 2014, 2016, 2019; Freire, 1970 y 2000, Vygotsky, entre muchos/as otros/as). El triángulo didáctico es un modelo que representa la relación esencial en el proceso de aprendizaje-enseñanza entre tres elementos: el profesor, el estudiante y el contenido. Este modelo se usa para describir las interacciones y el dinamismo que deben existir entre estos tres componentes para que el aprendizaje sea efectivo. Es decir, los tres componentes del fuertemente criticado triángulo didáctico no son suficientes para comprender y desarrollar crítica y transformadoramente los complejos procesos de aprendizaje, enseñanza, producción e investigación.

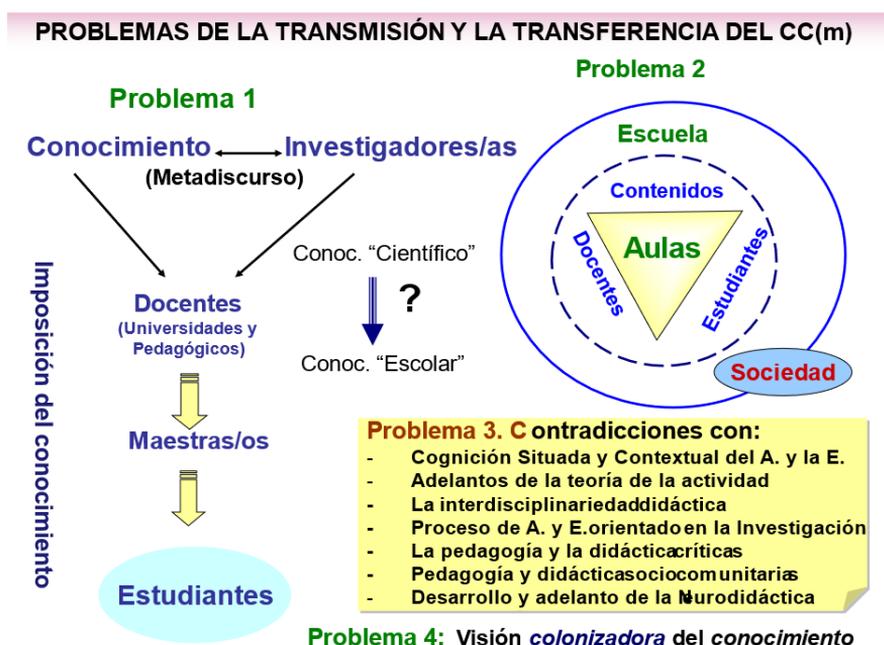
Los tres componentes del cuestionado triángulo didáctico, son presentados muy trivialmente, de la siguiente manera: a) El docente, quien tiene la

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

responsabilidad de transmitir conocimientos, orientar el aprendizaje y fomentar la participación activa de los estudiantes; b) El alumno (así lapidariamente tratado), quien recibe el conocimiento, se involucra en el proceso de aprendizaje, realiza preguntas y reflexiona sobre lo aprendido, pero sigue dependiendo como ente pasivo del proceso didáctico de los/as docentes, quienes tienen el poder y la palabra direccionadora de la enseñanza y el aprendizaje; y c) El contenido, el cual constituye el conocimiento que se va a enseñar los/as docentes con la ayuda de los materiales y libros de texto; el mismo debe estar estructurado y ser accesible para los estudiantes según sus necesidades y contexto.

Aquí entra a jugar un papel importante la denominada y conocida elementalización del conocimiento científico en conocimiento didáctico. En el esquema de la figura 8 mostramos los cuatro grandes problemas que caracterizan a la didáctica practicada en buena parte de las instituciones escolares en los diversos ámbitos de nuestros sistemas educativos.

Figura 8: Los 4 grandes problemas de la transmisión y transferencia del CC(m).



Fuente: Elaboración propia.

En muchas oportunidades hemos realizado fuertes y argumentativas críticas a esta conceptualización pedagógica y didáctica del triángulo didáctico, siempre desde el posicionamiento de la teoría crítica y, muy especialmente, desde las

consideraciones de la pedagogía y didácticas críticas (Mora, 2010, 2014, 2016, 2019).

Igualmente, variados pensadores y pedagogos han criticado este modelo por sus limitaciones en el enfoque tradicional de enseñanza y su tendencia a reducir el aprendizaje a una simple transmisión unidireccional del conocimiento, sin considerar aspectos fundamentales como la participación activa del estudiante, el tratamiento de temas generadores de aprendizaje y enseñanza, el contexto cultural y la crítica social.

Así, por ejemplo, Paulo Freire (1970 y 2000), cuestiona la visión tradicional de la educación representada en el triángulo didáctico, que considera al profesor como el único agente del conocimiento y al estudiante como objeto pasivo de enseñanza. En su obra *Pedagogía del oprimido* (1970), Freire introduce el concepto de educación dialógica, que plantea un proceso de aprendizaje-enseñanza basado en el diálogo entre los/as educadores/as y los/as educandos/as, donde ambos/as se enriquecen mutuamente.

El pedagogo brasileño es crítico a la educación bancaria, que se centra en la simple transmisión de informaciones de los/as profesores/as a los/as estudiantes, sin una verdadera participación ni crítica por parte los/as estudiantes. Según Freire, podríamos considerar que el triángulo didáctico no toma en cuenta las experiencias previas, el contexto socioeconómico y cultural de los/as estudiantes ni sus capacidades para construir saberes y conocimientos de manera activa, participativa, cooperativa y colaborativa.

Paulo Freire aborda su crítica a la educación tradicional y presenta su modelo pedagógico basado en el diálogo y la participación activa de todos/as los/as estudiantes, profundizando en la idea de que la educación debe ser liberadora y debe reconocer la autonomía de todos/as los/as que participan en la práctica pedagógica. La enseñanza debe ser más participativa, comprometida y crítica para generar conciencia sociopolítica y transformación social.

Por su parte, Wind y Mora, como otros críticos de la pedagogía tradicional, subrayan que el modelo del triángulo didáctico en su forma más básica mantiene una visión reduccionista del proceso de aprendizaje y enseñanza. Según él, la didáctica no debe centrarse sólo en la relación entre los/as docentes, el contenido y los/as estudiantes, sino que debe incluir los contextos socio-políticos, culturales y emocionales que influyen en el aprendizaje. Este modelo no contempla de manera adecuada la diversidad de los/as estudiantes ni sus experiencias previas, lo cual es esencial para un enfoque verdaderamente inclusivo, crítico y transformador. Wind y Mora abogan por una didáctica crítica que desafíe la imposición de un conocimiento único y homogéneo. Desde su perspectiva, el conocimiento no es neutral y está imbuido en relaciones de poder. El modelo tradicional del triángulo didáctico

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

puede ser visto como una herramienta de control social, cultural y cognitiva, donde los/as docentes ejercen poder sobre el saber, los conocimientos, el contenido y sobre los/as estudiantes.

El proceso de aprendizaje y enseñanza en este marco puede estar alineado con las lógicas de disciplina y normalización propias de las instituciones educativas, tal como lo han planteado en diversas obras Althusser, Gramsci y Ilich. Este análisis crítico que hacen Wind y Mora invita a repensar la educación, la formación, la pedagogía y la didáctica como un proceso en el que los/as estudiantes no sólo sean receptores/as pasivos/as, sino actores/as activos/as que cuestionan y transforman los saberes, los conocimientos y, sobre todo, las realidades en su máxima expresión. Otro de los/as grandes pensadores/as, especialmente en el campo de la psicología histórico cultural, es Lev Vygotsky.

Desde la perspectiva sociocultural de Vygotsky, el aprendizaje es social y situado. Él critica el enfoque aislado que podría tener planteamientos pedagógicos, tal como lo vemos en el triángulo didáctico, pues en su teoría el aprendizaje se da a través de la interacción con el entorno social y cultural. Para Vygotsky, la Zona de Desarrollo Próximo es clave, lo que implica que el conocimiento no se debe dar de manera unilateral ni direccional de unos/as que supuestamente lo saben todo (los/as docentes/as) a otros/as que también supuestamente carecen de luces parias, de autonomía y de saberes (los/as estudiantes), sino mediante la interacción, el apoyo social mutuo y el trabajo colaborativo compartido y altamente participativo entre estudiantes y docentes.

Es importante resaltar, además, los aportes de otros/as educadores/as críticos, como es el caso de Peter McLaren, quien se ha constituido en un referente fundamental de la pedagogía crítica contemporánea. El mencionado autor es crítico de la estructura de la enseñanza tradicional por reproducir relaciones de dominación y propone una educación orientada a la emancipación social. Para McLaren, los modelos como el triángulo didáctico son insuficientes si no se insertan en una perspectiva de transformación sociopolítica. Peter McLaren señala que los modelos didácticos clásicos no cuestionan la reproducción ideológica del sistema, por lo cual la relación profesor/a-alumno/a-contenido debe entenderse siempre dentro de un marco de lucha social y política. Igualmente, Henry Giroux profundiza en la necesidad de ver a los/as docentes no sólo como transmisores/as de conocimientos, sino como intelectuales transformadores/as que ayudan a los/as estudiantes a analizar críticamente la sociedad. El intelectual estadounidense es crítico con las concepciones reduccionistas de la didáctica que ven a los/as estudiantes como meros recipientes de saberes. Según Henry Giroux, el triángulo didáctico tradicional elimina el compromiso político del acto de educar y reduce la enseñanza a técnicas de transmisión, desconectadas de los procesos de formación crítica.

Por su parte Antonia Darder, influenciada directamente por Paulo Freire, plantea que todo proceso educativo es un acto político y debe reconocer las identidades culturales de los/as estudiantes. Ella hace un profundo análisis crítico de los modelos didácticos tradicionales por no incluir la dimensión cultural, afectiva y política del aprendizaje. Antonia Darder sostiene que la relación entre docente, contenido y estudiante debe estar mediada por el reconocimiento de las diferencias y la justicia social, aspectos omitidos en el triángulo didáctico convencional.

El triángulo didáctico, en su forma tradicional, ha sido criticado por reducir el aprendizaje a un proceso simple y lineal que no toma en cuenta la diversidad, el contexto, las necesidades emocionales y las capacidades de los/as estudiantes. Autores/as como Freire, Mora, Vygotsky McLaren, Giroux y Darder, entre muchos/as otros/as, proponen un modelo educativo más complejo y dinámico, donde los/as estudiantes son sujetos activos/as y el aprendizaje se construye en un contexto social, cultural, crítico y transformador. Para lograr un verdadero aprendizaje desde estas perspectivas, es necesario abandonar enfoques verticales y fomentar relaciones pedagógicas horizontales, dialógicas y cooperativas.

12. Las seis grandes dimensiones del hexágono pedagógico-didáctico de Wind y Mora

Tal como se muestra en la figura anterior, el triángulo pedagógico antes mencionado y descrito es incompleto e incompatible con una concepción más crítica y transformadora de la educación, la formación, la pedagogía y la didáctica. Eso lo podemos ver, desde la perspectiva hexagonal, en las tres grandes intervenciones: a) Primera dimensión, constituida por seis grandes categorías que determinan en buena medida el mundo de vida, constituida por la sociedad, la cultura, la economía, el ser humano, la naturaleza y la tecnología; b) La segunda dimensión, se refiere a las seis perspectivas que están directamente vinculadas con las últimas propuestas en torno la incorporación de la IAG en el mundo pedagógico, especialmente en el ámbito del Desarrollo de los Procesos de Aprendizaje y Enseñanza. En el esquema que aparece en la figura 9 podemos ver la interrelación hexagonal de las seis grandes dimensiones pedagógico-didáctica caracterizadoras de la nueva didáctica.

I. Perspectiva Tecnológica

Esta perspectiva se centra en cómo funciona una tecnología. En este nivel, se analizan los aspectos técnicos de la IA, como los componentes de la máquina, la arquitectura del software y los algoritmos. Con respecto a ChatGPT, por ejemplo, se puede señalar que este modelo está basado en GPT-3.5, una tecnología avanzada que funciona como un sistema de cubos con múltiples

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

caras. Cada vez que ChatGPT genera una respuesta, selecciona una cara del cubo en función de los datos de entrenamiento y los patrones estadísticos.

II. Perspectiva Sociocultural

Esta perspectiva examina cómo la tecnología impacta en la sociedad y las culturas. En el caso de la IA generativa, la OCDE (2023) resalta que las aplicaciones de la IA están afectando a un amplio rango de tareas y empleos, automatizando procesos rutinarios y creando una creciente demanda de habilidades específicas de IA en el mercado laboral. El impacto de tecnologías como ChatGPT está siendo cada vez más evidente en las áreas temáticas y los campos profesionales que la ZHAW cubre en sus programas de grado y formación continua.

III. Perspectiva Orientada a la Aplicación

Aquí, el enfoque se dirige a cómo se puede utilizar la tecnología en contextos específicos. Por ejemplo, en la educación, ChatGPT puede ser utilizado para traducir textos, resumir información o generar ideas para proyectos. Los estudiantes pueden beneficiarse de esta herramienta durante el proceso de escritura, como apoyo para la redacción o la reformulación de textos. Los profesores, por su parte, pueden emplear la IA para elaborar materiales didácticos o crear textos informativos.

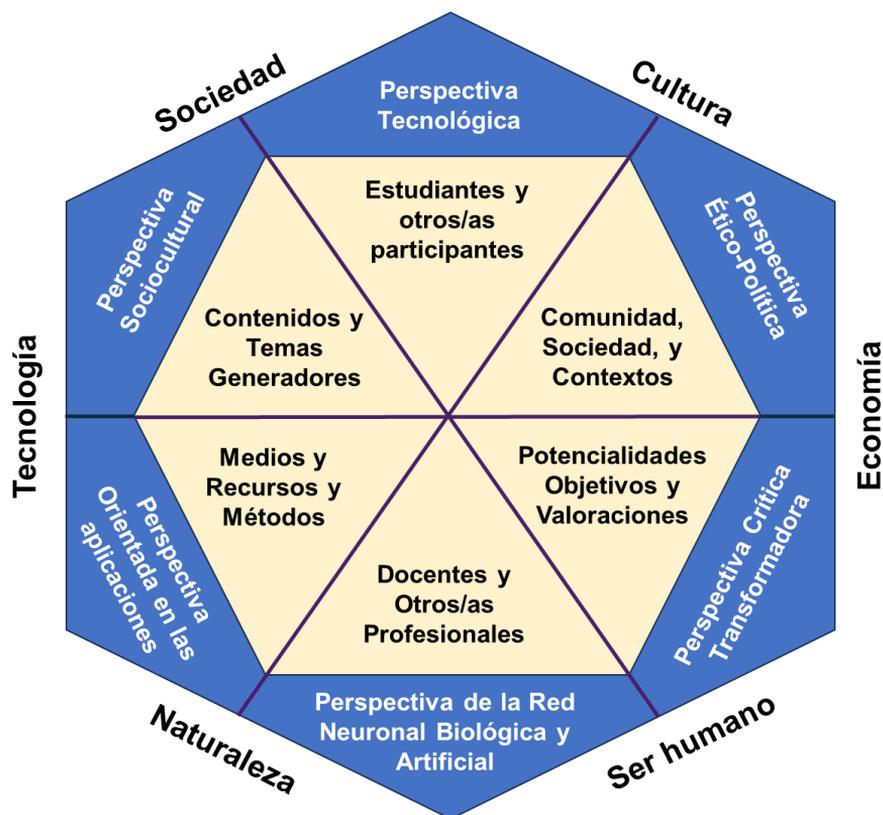
IV. Perspectiva crítica transformadora

Esta perspectiva se centra en cuestionar y reflexionar críticamente sobre las implicaciones éticas, sociales y pedagógicas de la IA generativa, con el objetivo de generar un cambio positivo y transformador. No se trata sólo de analizar cómo la tecnología funciona o se aplica, sino también de cuestionar las estructuras de poder, las dinámicas de exclusión y los impactos a largo plazo en la sociedad y la educación. En el contexto de la IA generativa en educación, esta perspectiva se enfoca en cómo la implementación de tecnologías como ChatGPT puede influir en la autonomía de los estudiantes, en la equidad en el acceso al conocimiento y en la relación entre docentes y estudiantes.

Además, plantea la necesidad de cuestionar el papel de los algoritmos y las empresas tecnológicas en la configuración de los contenidos y los procesos educativos. Por ejemplo, se puede analizar si el uso de plataformas de IA favorece la estandarización del aprendizaje o si, por el contrario, está deshumanizando las interacciones pedagógicas. La perspectiva crítica transformadora también plantea la importancia de diseñar políticas educativas y marcos regulatorios que aseguren que la tecnología no solo beneficie a aquellos que ya tienen acceso y poder, sino que sirva para reducir las

desigualdades en el sistema educativo. Esto implica no solo la integración de la IA en los currículos, sino también la formación ética de los educadores en el uso responsable de estas herramientas. Al mismo tiempo, subraya la necesidad de que los estudiantes desarrollen habilidades críticas para comprender y cuestionar los sesgos inherentes a los algoritmos y para navegar de manera reflexiva en un mundo digital cada vez más complejo.

Figura 9: Las seis grandes dimensiones del hexágono pedagógico de Wind y Mora



Fuente: Elaboración propia.

En resumen, la Perspectiva crítica transformadora propone una educación que no solo se basa en la adopción tecnológica, sino que también busca cuestionar y transformar las estructuras que afectan al aprendizaje y la sociedad, promoviendo un cambio hacia una práctica educativa más justa, equitativa y consciente de las implicaciones sociales y culturales de la tecnología.

V. Perspectiva Ético-Política

Esta perspectiva examina críticamente los dilemas éticos y políticos vinculados al uso de la IAG en contextos educativos. Se consideran fenómenos como el sesgo algorítmico, la vigilancia y la mercantilización de los datos personales, así como las estructuras de poder que las tecnologías de inteligencia artificial tienden a reproducir o, en algunos casos, a desafiar. La educación mediada por IAG no es neutral; está atravesada por decisiones éticas que afectan a quién se incluye, a quién se margina y qué formas de conocimiento se privilegian (Zuboff, 2019). En este sentido, el hexágono pedagógico pone especial énfasis en desarrollar un pensamiento ético y político tanto en estudiantes como en docentes, para fomentar prácticas pedagógicas que resistan la automatización acrítica y promuevan la justicia social en el ecosistema educativo (Williamson y Eynon, 2020).

VI. Perspectiva de la Red Neuronal Biológica y Artificial

Esta perspectiva se centra en la comparación entre las redes neuronales humanas (biológicas) y las redes neuronales artificiales que sustentan los sistemas de IAG. Aunque ambas comparten principios generales de procesamiento de información, existen diferencias fundamentales en cuanto a plasticidad, creatividad y conciencia, que es necesario comprender desde una mirada interdisciplinaria. Promover el conocimiento básico de neurociencia, informática y pedagogía permite desmitificar las capacidades de la IAG, subrayando que, pese a sus logros, sigue siendo un modelo estadístico de patrones, sin comprensión profunda o intención. Tal entendimiento es esencial para formar un alumnado crítico que pueda interactuar con la tecnología reconociendo sus alcances y limitaciones (Marcus y Davis, 2019). De este modo, la comparación no sólo tiene un valor técnico, sino también pedagógico y ético.

13. Hacia una pedagogía integral en tiempos de IA: Ocho aspectos complementarios del hexágono pedagógico de Wind y Mora

El hexágono pedagógico propuesto por Wind y Mora representa una superación conceptual del tradicional triángulo didáctico (docente, estudiante y contenido) al incorporar seis dimensiones fundamentales: la dimensión social, cultural, ecológica, humana, económica y tecnológica. Este modelo reconoce que el proceso educativo, especialmente en el contexto de la irrupción de la Inteligencia Artificial Generativa (IAG), no puede reducirse a la transmisión de saberes, sino que debe constituirse como una experiencia compleja, relacional y crítica. En complemento a estas seis dimensiones, pueden desarrollarse ocho aspectos articuladores y ampliados que fortalecen el alcance de este

hexágono y permiten una praxis pedagógica más integral: ético, comunicativo, formativo, cognitivo, emocional, intercultural, interdisciplinario y transformador.

En primer lugar, la dimensión *ética* no se reduce a la transmisión de normas morales, sino que implica una reflexión profunda sobre el sentido del acto educativo, la responsabilidad del/la educador/a, el respeto por la dignidad del/la estudiante y el juicio crítico frente a las tecnologías. La IA plantea dilemas inéditos —como la autoría de textos generados automáticamente, la equidad en el acceso digital o la vigilancia educativa— que exigen desarrollar en los sujetos una conciencia ética sólida, capaz de actuar con integridad en contextos complejos. Esta ética debe ser propositiva, situada y dialógica.

El aspecto *comunicativo*, por su parte, remite a la capacidad de construir significados compartidos a través del lenguaje verbal, no verbal y simbólico, en contextos mediados por tecnologías. En un entorno donde la interacción puede darse tanto cara a cara como mediante interfaces automatizadas, es fundamental fomentar habilidades de escucha activa, diálogo argumentado y comprensión intercultural. La comunicación pedagógica debe ser horizontal, empática y generadora de sentido, más aún cuando la IA tiende a reemplazar ciertas mediaciones humanas por algoritmos impersonales. La componente *formativa* integra todos los procesos orientados al desarrollo integral de la persona. No se limita a la instrucción técnica, sino que articula la formación de valores, actitudes, competencias y hábitos de pensamiento crítico y autónomo. En tiempos de IAG, la formación debe centrarse en aquello que las máquinas no pueden hacer: imaginar, interpretar, cuidar, deliberar éticamente. Una educación formativa debe propiciar el florecimiento humano, el compromiso social y la capacidad de actuar con responsabilidad en la transformación del entorno.

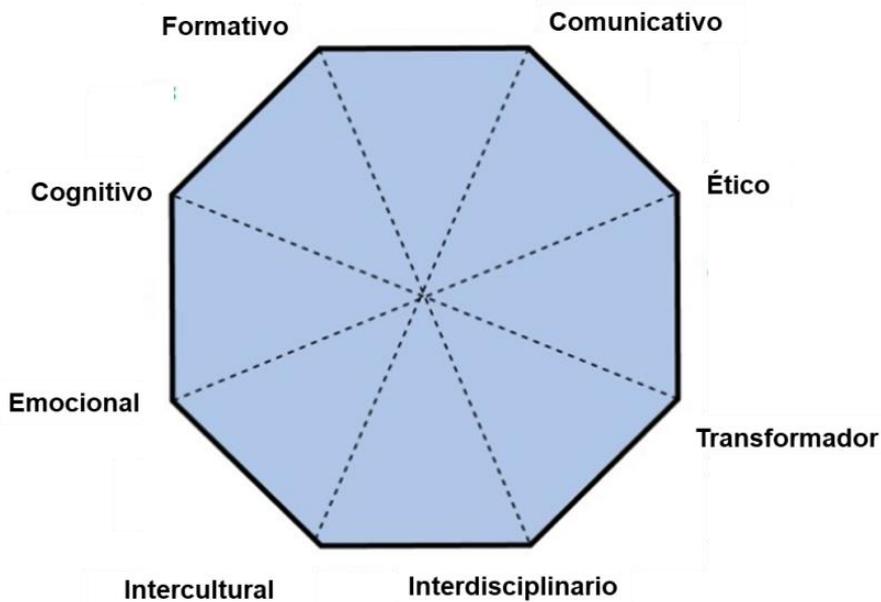
Desde el plano *cognitivo*, el proceso educativo se concibe como una construcción activa del conocimiento. Esto implica reconocer la diversidad de estilos y ritmos de aprendizaje, fomentar el pensamiento crítico y creativo, y diseñar ambientes de aprendizaje significativos. La IA ofrece múltiples recursos para potenciar los procesos cognitivos, pero su uso debe ser orientado por propósitos pedagógicos claros. No basta con „enseñar con IA“; se trata de enseñar a pensar críticamente sobre la IA y a pesar de la IA, recuperando la función reflexiva y problematizadora del acto educativo. La dimensión *emocional* reconoce que no hay aprendizaje auténtico sin involucramiento afectivo. Las emociones, lejos de ser distracciones, son elementos constitutivos del proceso de construcción de sentido. En la era de la automatización y la despersonalización digital, el/la educador/a debe cultivar un clima emocional que propicie la confianza, la motivación, el respeto y la resiliencia. La gestión emocional es clave para fortalecer la empatía, el trabajo colaborativo y la construcción de identidades sanas frente a los desafíos tecnológicos.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

El aspecto *intercultural* interpela la pedagogía a abrirse al reconocimiento y la valorización de la diversidad cultural, científica, epistemológica, lingüística, étnica y de género. En un mundo problemático, globalizado e interconectado, donde los algoritmos tienden a homogeneizar las experiencias y contenidos, en detrimento de la heterogeneidad cultural, es vital fomentar una pedagogía del reconocimiento, que promueva el diálogo entre saberes, la crítica al colonialismo epistémico y el derecho a aprender desde la propia cultura. La IA no puede ser neutra frente a estas tensiones: su desarrollo y uso también deben estar guiados por principios de justicia sociocultural.

El componente *interdisciplinario* complementa esta visión ampliada al promover una ruptura con la fragmentación del conocimiento. Frente a problemáticas complejas, como el cambio climático, la justicia algorítmica, la miseria, el hambre, la explotación, la discriminación o la ética del desarrollo tecnológico, se requiere una pedagogía capaz de articular múltiples disciplinas, enfoques y lenguajes, insistiendo en los elementos críticos integradores. El pensamiento sociocrítico, la colaboración entre campos y la creatividad teórico-práctica son potencialidades esenciales para formar sujetos capaces de afrontar los retos del presente desde múltiples miradas. En la figura 10 podemos apreciar una representación gráfica octagonal de estos ocho elementos complementarios e integradores.

Figura 10: Los ocho aspectos complementarios al hexágono pedagógico de Wind y Mora.



Fuente: Elaboración propia.

Finalmente, el aspecto *transformador* plantea que la educación y formación no pueden limitarse a la adaptación al sistema vigente, sino que deben orientarse hacia la emancipación personal y la transformación social. En este sentido, el hexágono pedagógico, complementado con el octágono crítico-transformador se proyecta como una herramienta para imaginar nuevas formas de aprender y enseñar, que superen la lógica de reproducción y permitan pensar críticamente el papel de la tecnología en nuestras vidas. Una pedagogía transformadora no se opone a la tecnología, pero tampoco se subordina a ella: la interroga, la contextualiza y la reconduce al servicio de fines humanos, democráticos y solidarios.

En suma, los ocho aspectos aquí desarrollados permiten enriquecer la propuesta del hexágono pedagógico con una mirada más holística, crítica y situada. En contextos marcados por la presencia creciente de la IA y la automatización educativa, el desafío no es sólo técnico, sino profundamente pedagógico: ¿cómo formar personas íntegras, sociocríticas y comprometidas en un mundo artificialmente inteligente? La respuesta, lejos de encontrarse en las máquinas, está en la capacidad de renovar la práctica educativa desde una ética del cuidado, una razón crítica y una esperanza transformadora.

14. Algunos de los desafíos principales de la integración de la IAG en los procesos didácticos

Aunque el uso de sistemas de IAG ya está siendo implementado en algunas asignaturas en algunos países, especialmente aquellos que apuestan por estas tecnologías en el mundo educativo y formativo, el mismo sigue siendo un fenómeno nuevo en muchos otros contextos educativos. Este proceso de integración requiere tiempo para consolidarse de manera que los/as estudiantes puedan aprender con IAG, mediante IAG y para la IAG. La calidad de los textos generados por la IAG es notable, pero es importante recordar que la IAG no tiene una comprensión real del contenido, lo que significa que los resultados generados no siempre son correctos. Por tanto, es esencial que los/as estudiantes no dependan exclusivamente de la IAG para el trabajo académico (Delorme, 2022), además de todas las implicaciones críticas sociales, cognitivas, culturales, etc., que representa la IAG para el ser humano en general y cada sujeto en particular.

Manfred Spitzer (2012, 2005, 2015a, 2015b, 2018, 2020, 2023, 2024, 2025, entre otros) sostiene que el uso de la inteligencia artificial y herramientas digitales en la educación no reemplaza los procesos cognitivos fundamentales ni la interacción social necesaria para un aprendizaje profundo, duradero y transformador. Más aún, advierte que confiar demasiado en sistemas automatizados puede empobrecer las habilidades críticas, motoras y sociales que se desarrollan con métodos educativos más tradicionales, muchos de los cuales inclusive no han sido realmente comprendidos o puestos en práctica

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

por los respectivos sistemas educativos en buena parte de los países del mundo.

Este reconocido neurocientífico y psiquiatra alemán, ha sido una voz amplia y profundamente crítica respecto al uso de la Inteligencia Artificial (IA) en términos generales, pero esencialmente en la educación. En su obra más reciente, „Künstliche Intelligenz: Dem Menschen überlegen – wie KI uns rettet und bedroht“ (2023), Spitzer analiza los impactos de la IA en diversos ámbitos, incluyendo la educación. Aunque reconoce los beneficios potenciales de la IA, advierte sobre riesgos como la manipulación, la pérdida de privacidad y la disminución de habilidades cognitivas debido a la dependencia tecnológica. Spitzer (2023, 148) señala, por ejemplo, que:

„La inteligencia artificial no reemplaza la inteligencia humana, especialmente en ámbitos como la educación, donde el aprendizaje emocional, social y experiencial es fundamental. La interacción humana es insustituible en el proceso educativo, ya que fomenta habilidades como la empatía, la creatividad y el pensamiento crítico, que las máquinas no pueden replicar“.

Además de esta publicación, Spitzer ha abordado temas relacionados en obras anteriores, como „Digitale Demenz“ (2012), donde argumenta que el uso excesivo de medios digitales puede afectar negativamente el desarrollo cognitivo de los jóvenes. En „Cyberkrank!“ (2015), explora cómo la digitalización puede influir en la salud mental y física. En „Die Smartphone-Epidemie“ (2018), examina el impacto de los teléfonos inteligentes en la concentración y el aprendizaje. Y en „Künstliche Intelligenz und Bildung“ (2020), discute específicamente la relación entre la IA y la educación.

Todas estas obras reflejan la preocupación de Spitzer por los efectos de la tecnología en la educación y el desarrollo de los jóvenes, ofreciendo una perspectiva crítica sobre la integración de la IA en entornos educativos, tal como lo resume en uno de sus últimos libros (Spitzer, 2023, 201):

„Quien introduce tecnologías inteligentes demasiado pronto en la vida de los niños pone en peligro su desarrollo cognitivo, social y emocional. La exposición temprana y excesiva a dispositivos digitales puede llevar a una disminución en la capacidad de concentración, problemas de aprendizaje y una reducción en las habilidades interpersonales esenciales para el desarrollo integral“.

Spitzer (2012) introduce el concepto de „demencia digital“, argumentando que el uso prematuro y excesivo de tecnologías digitales puede atrofiar el cerebro de los/as niños/as. Aunque no se centra exclusivamente en IA, es clave para entender su postura general sobre la digitalización en la educación, lo cual ha sostenido desde hace más de veinte años. Sus argumentos se centran en

cómo estas tecnologías pueden perjudicar el desarrollo cognitivo, la memoria y las habilidades sociales, siendo algunos de ellos los siguientes:

-El uso excesivo de herramientas digitales en las aulas (como ordenadores, tablets o pizarras digitales) reduce la profundidad del procesamiento mental, esencial para el aprendizaje significativo.

-Externalización del pensamiento: La dependencia de IA y dispositivos digitales lleva a los estudiantes a delegar tareas cognitivas como, por ejemplo, la memorización o la orientación espacial, lo que atrofia habilidades cerebrales, el uso de GPS debilita la capacidad de orientación.

-Aprendizaje superficial: La multitarea digital (navegación rápida, hipervínculos) fragmenta la atención y dificulta la retención de información; él señala que leer en papel es más efectivo que en pantallas.

-Riesgos psicosociales: Asocia el uso temprano de IA/digitales con mayores tasas de adicción, depresión, déficit de atención y aislamiento social, citando casos como Corea del Sur, donde el 12% de los escolares eran adictos a internet en 2010. Él considera que es necesario insistir en la educación analógica, priorizar lápiz, papel, lectura profunda e interacción humana, lo cual requiere la prohibición de dispositivos digitales en aulas hasta la adultez temprana, hasta los 16 o 18 años de edad.

En definitiva, Spitzer (2023) expone con claridad tanto las ventajas como las desventajas de la inteligencia artificial, subrayando que estas últimas no radican en la tecnología en sí, sino en los usos que de ella se hacen. Si los escolares y aprendices se limitan a recurrir a herramientas como ChatGPT, descuidando el desarrollo del pensamiento autónomo y las interacciones reales con otras personas, el deterioro cultural que ya se evidencia con la adicción a los teléfonos móviles y las redes sociales podría agravarse de forma aún más profunda. Asimismo, la aplicación militar de la IA representa un factor de incertidumbre absoluta, especialmente cuando se encuentra en manos de líderes políticos atrapados entre decisiones extremas. En este escenario, donde se conjugan amenazas militares con intentos de predicción de catástrofes naturales, la inteligencia artificial se revela como un arma de doble filo. A ello se suma el ensanchamiento de la brecha social: mientras los beneficios se concentran en bolsillos privados, la carga financiera recae sobre el conjunto de la sociedad, que, a su vez, se vuelve cada vez más susceptible a la manipulación.

Por otra parte, otros autores/as también analizan esta problemática, tomando en cuenta los rasantes avances de la IAG y su masivo uso en diversos campos de la vida actual. Así, por ejemplo, Cameron et al. (2021) señalan que, aunque IAG puede ser de gran ayuda para la escritura, los/as estudiantes

deben entender que la escritura es mucho más que simplemente generar un texto. La escritura es una herramienta clave para el pensamiento, el aprendizaje y la comunicación. A través de la escritura, los/as estudiantes aprenden a organizar y argumentar sus pensamientos en un lenguaje académico apropiado para su campo. Este proceso sigue siendo fundamental para las carreras científicas y profesionales.

15. La descualificación como una consecuencia negativa de la IAG

El riesgo de pérdida de competencias no es exclusivo de la inteligencia artificial; este fenómeno ha acompañado históricamente el uso de herramientas y medios técnicos que las personas emplean para reducir su carga de esfuerzo físico o cognitivo. Desde los primeros análisis sobre los efectos de la industrialización, se ha observado un proceso de descualificación vinculado a la separación entre diseño y ejecución del trabajo, lo que implicó una transformación estructural en la naturaleza de las tareas laborales (Braverman, 1974). La denominada „deconstrucción de la artesanía“ constituyó un punto de inflexión al reemplazar habilidades complejas y contextuales por operaciones estandarizadas, lo cual sentó las bases para una racionalización del trabajo que sigue vigente en la actualidad (Sennett, 2008).

Los ejemplos iniciales de descualificación en la era industrial revelan un patrón recurrente: unas pocas personas con alta calificación diseñan, operan y supervisan las máquinas, mientras que el grueso de la fuerza laboral queda relegado a funciones repetitivas o de bajo nivel técnico. Esta brecha entre trabajos poco cualificados y aquellos altamente especializados se ha ampliado con la digitalización, aunque en sus primeras fases también permitió una mejora general del capital humano mediante la demanda de nuevas competencias cognitivas, no rutinarias y más creativas (Autor, Levy y Murnane, 2003). La tecnología, en estos casos, no sólo sustituyó funciones, sino que abrió espacios para el desarrollo de habilidades de orden superior, demostrando un potencial ambivalente de erosión y enriquecimiento profesional.

Sin embargo, no existe consenso sobre si esta tendencia histórica se mantendrá frente a la expansión de las tecnologías basadas en inteligencia artificial. Mientras algunos autores sostienen que las olas tecnológicas anteriores ofrecen un marco de continuidad —donde la pérdida de ciertos empleos puede compensarse con la creación de otros más sofisticados—, otros advierten que la IA generativa introduce un nivel de complejidad sin precedentes, capaz de afectar transversalmente a todos los sectores, incluyendo aquellos tradicionalmente protegidos por el conocimiento especializado (Brynjolfsson y McAfee, 2014; Susskind, 2020a). En este escenario, el trabajador del conocimiento ya no sería el beneficiario principal

del progreso tecnológico, sino un actor potencialmente vulnerable a una nueva ola de descualificación. Esta incertidumbre refleja, en gran medida, la rapidez y la imprevisibilidad del desarrollo actual de la IA, lo cual justifica la necesidad de examinar más detenidamente su conceptualización y sus implicaciones.

El término IAG y su relación con la sobrecarga de trabajo

A menudo se parte del supuesto de que el concepto de inteligencia artificial (IA) está claramente definido. No obstante, un examen más riguroso revela una multiplicidad de interpretaciones, enfoques y usos del término. En sentido estricto, la IA constituye una categoría amplia que agrupa distintas herramientas informáticas diseñadas para resolver tareas específicas mediante el procesamiento sistemático de información. Así entendida, y dejando de lado las interpretaciones metafóricas, „inteligencia artificial“ puede definirse como un conjunto de procesos técnicos orientados a la gestión y transformación de datos (Russell y Norvig, 2021).

Desde esta perspectiva, el término „artificial“ alude a lo técnico, mientras que «inteligencia» remite al procesamiento de información, lo cual difiere sustancialmente de las atribuciones antropomórficas que se suelen proyectar sobre las máquinas. Diversos autores han advertido sobre el carácter metafórico de esta denominación, subrayando que el concepto de inteligencia humana no puede extrapolarse de manera directa a sistemas técnicos (Floridi, 2022). En el ámbito de la psicología, por ejemplo, la inteligencia se conceptualiza como un constructo hipotético e inobservable directamente, definido por la capacidad de comprensión, juicio, razonamiento, pensamiento orientado a objetivos y adaptación eficaz al entorno (Sternberg, 1985). Esta distancia conceptual entre inteligencia humana e inteligencia artificial requiere una clasificación más matizada del desarrollo técnico.

En este sentido, algunos autores han propuesto tipologías que permiten ubicar la IA en el marco histórico-cultural de la evolución de los medios técnicos. Estas clasificaciones distinguen etapas que van desde: (1) la herramienta, donde todos los aspectos de la acción técnica surgen del cuerpo humano; (2) la máquina, que externaliza la generación de energía; (3) el autómat, que automatiza movimientos físicos repetitivos; (4) el ordenador, capaz de ejecutar operaciones intelectuales sistemáticas; (5) la IA débil, que resuelve problemas mediante algoritmos adaptativos; (6) la autonomía técnicamente incorporada, que interviene en procesos físicos de resolución de problemas; (7) la semiautonomía técnica, que incluye la capacidad de fijar objetivos limitados; hasta llegar a (8) la IA fuerte o autonomía técnica plena, que implicaría autodefinición de metas y control autónomo, aunque esta última aún permanece en el plano de lo hipotético (Kaplan, 2016; Tegmark, 2017).

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

Estas fases no sólo trazan una cronología del progreso tecnológico, sino que también permiten anticipar distintas consecuencias en términos de cualificación laboral. En efecto, dependiendo de si la tecnología sustituye rutinas físicas, tareas cognitivas repetitivas o incluso procesos intelectuales complejos, pueden producirse diversas formas de descualificación o, en contrapartida, nuevas exigencias formativas. Por lo tanto, desde una perspectiva teórica, cabe plantearse que la IA generativa —dada su capacidad de abordar tareas no estructuradas y creativas— podría constituir una inflexión en la tendencia histórica de alternancia entre pérdida y fortalecimiento de habilidades, afectando incluso a las actividades propias del conocimiento humano.

Elementos caracterizadores de la pérdida potencial de cualificaciones humanas

Desde una perspectiva eminentemente económica, podría parecer lógico cuestionar si la pérdida de competencias —*deskilling*— representa realmente un problema. Si la inteligencia artificial (IA) se consolida como una presencia ubicua en todos los ámbitos sociales y laborales, operando de forma más eficiente y eficaz que los seres humanos, cabría suponer que las competencias reemplazadas dejarían de ser necesarias, y por tanto, no sería preciso conservarlas ni desarrollarlas nuevamente.

Sin embargo, esta conclusión resulta insatisfactoria desde un enfoque más amplio. El valor de una competencia no se limita a su utilidad económica inmediata, sino que depende también del papel que desempeña en el trabajo, la vida social y la constitución del sujeto humano. Si se considera esta dimensión, la pérdida de competencias puede implicar no sólo riesgos económicos, sino también impactos negativos en la estructura social y en el desarrollo personal (Eubanks, 2018; German Ethics Council, 2023). En los debates recientes sobre la relación entre IA y *deskilling*, como los recogidos por el Deutscher Ethikrat (Consejo Alemán de Ética), pueden identificarse al menos tres argumentos centrales que justifican una atención crítica al fenómeno. (a) En primer lugar, el crecimiento de la dependencia tecnológica en sectores cada vez más amplios y críticos de la sociedad no garantiza que la IA funcione sin fallos o esté siempre disponible.

Ante un fallo técnico o una interrupción en el servicio, la sociedad vuelve a depender de las capacidades humanas, que podrían haberse atrofiado por falta de uso. En este contexto, la pérdida de competencias humanas representa un riesgo sistémico. (b) En segundo lugar, mientras los seres humanos estén encargados de supervisar la IA y de intervenir en situaciones imprevistas, será indispensable que comprendan a fondo los procesos delegados. Sin embargo, este conocimiento se debilita si las personas ya no practican esas actividades de forma regular, debido a la automatización de las

rutinas (Pasquale, 2020). La paradoja consiste en que cuanto más eficiente es la IA, menor es la experiencia humana acumulada para actuar cuando falla. (c) Por último, desde una perspectiva antropológica, el *deskilling* también puede afectar la autonomía y la autopercepción del individuo. La delegación constante de decisiones a sistemas técnicos podría debilitar el sentido de agencia personal, y con ello, disminuir la participación activa en la vida social y cívica (Zuboff, 2019).

A esta última preocupación cabe añadir una dimensión relacional: la sustitución progresiva de interacciones humanas por vínculos con sistemas de IA —especialmente aquellos basados en modelos generativos del lenguaje— puede empobrecer el tejido social. Si el diálogo, la colaboración, la retroalimentación y el apoyo mutuo se sustituyen por relaciones funcionales con sistemas artificiales, se reduce la práctica de habilidades interpersonales fundamentales como la empatía, la paciencia o la disposición al compromiso (Turkle, 2015). De este modo, la creciente mediación tecnológica no sólo redefine los procesos de trabajo, sino que también transforma las expectativas comunicativas y afecta la calidad de las relaciones humanas en el largo plazo.

La pérdida de las cualificaciones científicas, académicas y prácticas de lo/as docentes universitarios/as

La irrupción pública de ChatGPT a finales de 2022 marcó un punto de inflexión significativa, no solo en el debate sobre inteligencia artificial (IA), sino también en la percepción generalizada de su alcance, especialmente dentro del ámbito académico no especializado en ingeniería o informática. Por primera vez, se hizo evidente que la IA podría aplicarse con eficacia en dominios tradicionalmente considerados patrimonio exclusivo de las capacidades humanas, tales como la creatividad intelectual, la formulación de ideas, la resolución de problemas complejos, el diseño metodológico y la escritura académica.

Estas actividades, que conforman lo que algunos autores denominan *trabajo del conocimiento*, son precisamente las que han sido objeto de reflexión en torno a los efectos de la IA tanto en términos de mejora como de descalificación (Drucker, 1999; Susskind, 2020b).

En un tiempo notablemente breve, numerosos investigadores, docentes y estudiantes comenzaron a experimentar con las posibilidades que ofrece ChatGPT como modelo lingüístico de uso general en entornos universitarios. Existe cierto consenso en que esta tecnología transformará el quehacer académico, y muchos consideran que se inscribe en la línea evolutiva de innovaciones digitales previas. Desde esta perspectiva, puede hablarse de un renovado interés por las oportunidades que ofrece la IA generativa para el fortalecimiento de competencias. Se sostiene que herramientas como

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

ChatGPT permiten liberar tiempo para dedicarse a tareas cognitivas más complejas y de mayor valor añadido, en una lógica ya clásica de la sociología del trabajo: al automatizar procesos rutinarios, se incentiva la reasignación hacia funciones más exigentes que potencian la cualificación (Autor, 2015). Sin embargo, la reflexión sobre los riesgos de descualificación en la educación superior ha sido escasa y poco desarrollada, tanto en el plano conceptual como empírico.

Uno de los pocos espacios en los que se aborda indirectamente este fenómeno es el de la evaluación: el uso de ChatGPT ha generado preocupación respecto a la validez de las pruebas escritas, como los ensayos y tareas domiciliarias. Se teme que el uso indebido de IA pueda comprometer la autenticidad del rendimiento académico, impidiendo discernir si el estudiante conserva ciertas competencias fundamentales o si estas han sido sustituidas sin oportunidad de ejercitación (Cotton et al., 2023). Esta problemática revela una preocupación importante: el foco suele ponerse en la verificación de conocimientos adquiridos, pero menos en la posibilidad de que dichas competencias ni siquiera se desarrollen cuando se delegan en exceso en la tecnología.

En el ámbito de la enseñanza, particularmente en la escritura académica, las discusiones sobre la posible pérdida de capacidades son más frecuentes. Diversos expertos en didáctica de la escritura investigan actualmente cómo cambia la producción textual cuando se integra el uso de IA generativa. No solo se identifican nuevas exigencias relacionadas con el uso ético y competente de estas herramientas, sino también riesgos concretos como la disminución de la originalidad, la pérdida de voz autoral y el debilitamiento del pensamiento crítico al relegar los procesos reflexivos a una máquina (Re et al., 2023; Bender et al., 2021).

En cuanto a la investigación científica, las herramientas digitales —y en particular la IA generativa— están comenzando a desempeñar un papel relevante en múltiples fases del proceso investigador, desde la búsqueda y análisis de datos hasta la formulación de hipótesis o la redacción de resultados. Si bien se reconoce su utilidad para asistir a los científicos en tareas complejas, también se advierte sobre consecuencias negativas como la homogeneización del conocimiento, la disminución de la diversidad epistémica y la erosión de la integridad científica (Heaven, 2023). Solo algunos autores, sin embargo, abordan el riesgo de descualificación de forma explícita. Por ejemplo, Björn Brembs (2023) advierte sobre la posibilidad de que se genere una „indefensión aprendida“, en la que se pierda progresivamente la capacidad de construir teorías y de comprender el mundo social de forma independiente, como resultado del uso excesivo y acrítico de sistemas como ChatGPT en las ciencias humanas y sociales.

Conclusión crítica: Inteligencia artificial, competencias humanas y el riesgo de descualificación

La irrupción de la inteligencia artificial generativa en la vida social y académica ha reconfigurado radicalmente los marcos tradicionales de producción de conocimiento, ejecución técnica y desarrollo de competencias. A pesar de la apariencia de claridad conceptual, el término „inteligencia artificial“ encierra ambigüedades semánticas y metáforas que suelen encubrir su verdadera naturaleza: un procesamiento técnico automatizado de información que, si bien simula comportamientos inteligentes, dista mucho de emular la complejidad cognitiva y emocional del ser humano.

Desde una perspectiva histórico-técnica, la evolución de la IA puede interpretarse como parte de una serie de transiciones hacia la delegación progresiva de funciones humanas —primero físicas y luego cognitivas— en dispositivos tecnológicos. Esta evolución ha venido acompañada tanto de procesos de mejora de cualificaciones como de fenómenos de „deskilling“, o pérdida de capacidades, cuyos efectos son aún insuficientemente comprendidos. La clave del problema no radica simplemente en que ciertas tareas sean automatizadas, sino en qué tipo de competencias se ven afectadas y cuál es su valor para la autonomía individual, la cohesión social y el funcionamiento democrático.

El debate contemporáneo ha tendido a centrarse en los beneficios funcionales de la IA, eficiencia, ahorro de tiempo, personalización, minimizando o ignorando los riesgos estructurales asociados a la pérdida de habilidades humanas esenciales. Esta omisión es particularmente preocupante en contextos como la educación superior, donde las funciones centrales del trabajo académico, pensar, escribir, analizar, decidir, corren el riesgo de ser tercerizadas a sistemas opacos, debilitando potencialmente la agencia crítica de estudiantes e investigadores. Si la escritura, la resolución de problemas o la formulación de teorías son progresivamente delegadas a la IA, la consecuencia no es solo una redefinición del rol humano, sino la posible atrofia de las competencias mismas que sostienen la producción autónoma de conocimiento.

Más aún, el riesgo no se limita a lo individual. Una sociedad que transfiere a las máquinas tareas fundamentales de comunicación, deliberación y toma de decisiones se expone a una forma de dependencia estructural que podría comprometer su resiliencia ante fallas tecnológicas o crisis epistémicas. Como advierte el Deutscher Ethikrat, mantener ciertas capacidades humanas activas y disponibles no es solo deseable, sino necesario para la seguridad, la soberanía y la dignidad humanas.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

En síntesis, el fenómeno de la descualificación no puede ser interpretado únicamente como un problema técnico o económico, sino como un desafío antropológico y ético. Requiere de respuestas institucionales, educativas y culturales que promuevan un uso consciente, limitado y contextual de la inteligencia artificial, evitando su fetichización como solución universal. La pregunta central ya no es si debemos usar IA, sino cómo lo hacemos sin renunciar a lo que nos hace humanos.

¿Qué caracteriza, distingue y constituye la inteligencia humana en relación con la IAG?

En el campo de la psicología, la inteligencia humana se define como la capacidad para aprender de la experiencia, resolver problemas y adaptarse a situaciones nuevas (Neisser et al., 1996). Esta definición incorpora dimensiones múltiples, entre las que se encuentran las competencias lingüísticas, las habilidades lógico-formales, la imaginación espacial, la creatividad, la motricidad fina, así como las capacidades sociales y emocionales. Existe un debate persistente respecto a si estas dimensiones constituyen inteligencias independientes —como sugiere Gardner (1983) en su teoría de las inteligencias múltiples— o si dependen de una inteligencia general subyacente, comúnmente denominada „factor g“ (Spearman, 1961).

La medición de la inteligencia se ha llevado a cabo tradicionalmente mediante tests psicométricos diseñados para evaluar el cociente intelectual (CI), el cual se calcula como un valor numérico derivado del desempeño en tareas específicas. Muchos de estos tests incluyen ejercicios como secuencias numéricas, que requieren detectar patrones regulares para predecir el siguiente elemento de la serie. Algunos desarrollos en inteligencia artificial (IA) han logrado replicar este tipo de tareas con una precisión incluso superior a la humana. No obstante, estas capacidades se limitan a contextos específicos y no implican una inteligencia general: los sistemas pueden resolver una tarea delimitada sin poder generalizar ese conocimiento a otros dominios (Lake et al., 2017).

A menudo se comete el error de atribuir a los sistemas de IA capacidades humanas completas debido a la antropomorfización, es decir, la tendencia a proyectar características humanas en entidades no humanas. Por ejemplo, cuando se informa que una IA supera a dermatólogos humanos en el diagnóstico de cáncer de piel, se asume erróneamente que este sistema también posee comprensión médica general, habilidades clínicas o juicio moral. Este fenómeno se repite en el ámbito de los vehículos autónomos: se presume que la IA reconoce a los peatones como lo haría un humano, evaluando género, edad o etnicidad, cuando en realidad se basa en patrones estadísticos de forma y color (Marcus y Davis, 2019).

Una de las razones fundamentales por las cuales la IA actual no posee inteligencia general se relaciona con la ausencia de conciencia. Aunque existen múltiples enfoques filosóficos y científicos sobre qué constituye la conciencia, suele entenderse como la capacidad de experimentar estados mentales internos y de desarrollar autoconciencia (Block, 1995; Chalmers, 1996). La experiencia subjetiva, como la percepción de un color, no puede ser replicada por un sistema que simplemente etiqueta patrones de datos. Así, mientras una IA puede etiquetar correctamente una imagen como „flor azul“, no tiene una experiencia sensorial asociada a dicha percepción, como sí la tiene un ser humano.

Otros componentes esenciales de la inteligencia humana incluyen la fijación de metas a corto y largo plazo, la incorporación del conocimiento previo (conocido como sentido común), y la adaptación contextual del comportamiento. Este conocimiento intuitivo, como el entendimiento de que una pirámide no puede sostener un cuboide en una torre estable, es aprendido sin instrucción formal y representa una de las barreras más difíciles de superar para la IA (Davis y Marcus, 2015).

Asimismo, la inteligencia humana se caracteriza por la comprensión comunicativa. En el diálogo, los humanos infieren significados implícitos, detectan malentendidos y ajustan su comunicación en consecuencia. Los sistemas de procesamiento del lenguaje natural actuales —como GPT-3— son capaces de generar respuestas coherentes, pero carecen de comprensión semántica profunda. Por ejemplo, pueden identificar literalmente quién salió a pasear en un texto, pero tienen dificultades para inferir que un personaje vio un gato si esa información requiere inferencia lógica a partir de varias frases (Marcus y Davis, 2019).

En conclusión, aunque los avances en IA han alcanzado niveles notables en tareas específicas, la inteligencia general humana sigue siendo inimitable debido a su carácter holístico, experiencial, contextual y consciente. La investigación actual en IA general (AGI, por sus siglas en inglés) enfrenta no solo desafíos técnicos, sino también epistemológicos y filosóficos, que requieren una comprensión más profunda de lo que implica realmente ser inteligente.

16. Aprendizaje Humano frente a Aprendizaje Automático

El estudio del aprendizaje humano frente al aprendizaje automático (también denominado *machine learning*) ha sido un área clave de investigación en los campos de la psicología cognitiva y la inteligencia artificial. En muchas ocasiones, tratamos de trasladar las características del aprendizaje humano a los sistemas de inteligencia artificial, aunque existen diferencias fundamentales que marcan la distinción entre ambos procesos. A diferencia

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

del aprendizaje humano, que se basa en una integración constante de datos y conocimientos previos adquiridos, los modelos de aprendizaje automático suelen ser diseñados para funcionar con grandes cantidades de datos específicos, sin la incorporación de la experiencia adquirida a lo largo del tiempo (Chalmers, 1996; Davis y Marcus, 2015). El aprendizaje humano no solo se basa en la acumulación de datos, sino que incorpora el conocimiento previamente adquirido. Este proceso de aprendizaje es un fenómeno continuo que dura toda la vida, lo que permite a los seres humanos adaptarse a nuevas situaciones y contextos mediante la integración de experiencias pasadas y habilidades cognitivas generalizables (Gardner, 1983). En cambio, la mayoría de los enfoques de aprendizaje automático (a excepción de algunos enfoques como el aprendizaje por refuerzo) se basan exclusivamente en datos y en la creación de un modelo que, una vez entrenado, rara vez es modificado a menos que se realicen nuevos entrenamientos. Este aspecto es crucial para comprender la rigidez de los sistemas de IA frente a la flexibilidad inherente al aprendizaje humano (Lake, Ullman, Tenenbaum y Gershman, 2017).

Para ilustrar esta diferencia, tomemos el ejemplo del entrenamiento de una red neuronal para diferenciar entre imágenes de perros y gatos. En este caso, el modelo entrenado reconocerá con precisión los perros y gatos en nuevas imágenes basadas en las características de los ejemplos previamente mostrados. Sin embargo, si se introduce una imagen de un objeto completamente diferente, como un muffin o incluso un frigorífico, el modelo podría clasificar erróneamente la imagen como un perro o un gato, ya que el modelo no tiene la capacidad de reconocer lo „desconocido“ fuera del contexto de los datos en los que fue entrenado. En cambio, un ser humano, al observar la misma imagen, podría reconocer la novedad del objeto y clasificarlo como algo que aún no ha aprendido, evidenciando la flexibilidad del razonamiento humano frente a la rigidez de la IA (Block, 1995; Marcus y Davis, 2019).

El aprendizaje humano, por lo tanto, es inherentemente más eficiente en términos de datos, dado que los humanos pueden aprender de pocos ejemplos y generalizar de manera más efectiva. Un claro ejemplo de esto es el proceso de aprendizaje de la conjugación de verbos en inglés. Los niños, al aprender el pasado de los verbos regulares, generalizan la regla del sufijo „-ed“ a partir de unos pocos ejemplos, como *watch/watched* y *wait/waited*. Sin embargo, también pueden cometer errores al aplicar esta regla a verbos irregulares, como cambiar *win* por *wined* en lugar de *won* (Spearman, 1961). Este fenómeno subraya la capacidad del ser humano para reconocer regularidades y generalizar, lo que constituye la base del razonamiento inductivo. En el razonamiento inductivo, los individuos extraen una regla general a partir de ejemplos específicos, mientras que, en el razonamiento deductivo, se parte de premisas generales para llegar a conclusiones más específicas. A diferencia del razonamiento deductivo, la conclusión inductiva no siempre garantiza la certeza del resultado, ya que el conocimiento sigue

siendo imperfecto y sujeto a la revisión a medida que surgen nuevas experiencias (Neisser et al., 1996).

El conocimiento previo es un componente esencial del aprendizaje humano. A lo largo de nuestra vida, adquirimos conocimientos y habilidades específicas que nos permiten aprender conceptos más complejos y generales de manera más eficiente. Durante la infancia, los niños adquieren conceptos concretos y visuales junto con sus respectivas designaciones lingüísticas. Estos aprendizajes iniciales, aunque parecen simples y naturales, son fundamentales para el desarrollo cognitivo (Gardner, 1983). Sin embargo, al intentar reproducir este proceso de aprendizaje conceptual en sistemas de inteligencia artificial, se hace evidente la complejidad y flexibilidad del aprendizaje humano. Los científicos cognitivos han ilustrado de manera significativa esta diferencia con ejemplos artificiales. Un concepto como el de „toba“, un tipo de formación geológica, se aprende de manera intuitiva observando ejemplos variados. Aunque los ejemplos de tobas pueden presentar diferencias, los seres humanos son capaces de abstraer los elementos esenciales de los mismos, clasificando correctamente las imágenes según lo aprendido. Sin embargo, para que una red neuronal convolucional (CNN) aprenda este concepto, necesitaría una cantidad considerablemente mayor de ejemplos etiquetados y entrenados, lo que resalta las limitaciones de los sistemas de IA en cuanto a su capacidad para generalizar a partir de pocos datos (Lake et al., 2017; Marcus y Davis, 2019). Este fenómeno refleja las diferencias fundamentales en los procesos de aprendizaje entre humanos y máquinas.

En definitiva, las diferencias entre el aprendizaje humano y el aprendizaje automático son claras y profundas. Mientras que el aprendizaje humano se caracteriza por su flexibilidad, capacidad de generalización y dependencia del conocimiento previo, el aprendizaje automático depende principalmente de la acumulación de datos y la construcción de modelos estáticos. Aunque los avances en inteligencia artificial han sido impresionantes, los sistemas actuales siguen siendo limitados en su capacidad para adaptarse de manera similar a los humanos frente a nuevas situaciones o conceptos. Estas diferencias subrayan la importancia de continuar investigando en el desarrollo de sistemas de IA más flexibles y eficientes, capaces de emular, de manera más precisa, la capacidad humana para aprender y adaptarse (Chalmers, 1996; Marcus y Davis, 2019).

17. Conclusiones

El análisis detallado de los veintidós ejes temáticos abordados en este documento permite establecer una conclusión central: la Inteligencia Artificial (IA), y en especial la Inteligencia Artificial Generativa (IAG), no puede ser comprendida ni gestionada exclusivamente como una herramienta técnica.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

Más bien, constituye un fenómeno estructurante de la cultura contemporánea, capaz de reconfigurar las lógicas de producción, comunicación, aprendizaje, trabajo y subjetivación. En ese marco, su incorporación en los distintos ámbitos del mundo de vida debe ser evaluada no sólo por su eficiencia funcional, sino también por su impacto ético, educativo, formativo, cognitivo y social. El recorrido propuesto muestra que los orígenes de la IA están profundamente ligados a la historia del pensamiento lógico-matemático. Figuras como Alan Turing, influenciado por Gödel, Hilbert y Boole, no sólo sentaron las bases de la computación moderna, sino que también introdujeron interrogantes que siguen vigentes: ¿pueden las máquinas pensar? ¿Puede simularse la inteligencia? ¿Qué distingue a una operación mecánica de un proceso de comprensión? La crítica filosófica de Searle y su célebre „Habitación China“ aportan un contrapunto necesario a la fascinación tecnológica, recordando que el procesamiento sintáctico no equivale a la comprensión semántica, ni mucho menos a la conciencia o la intención.

Desde el punto de vista técnico, el surgimiento de las redes neuronales y la modelación computacional de la neurona biológica representan avances notables en la simulación de funciones cognitivas humanas. Aplicaciones como el diagnóstico médico asistido por IA dan cuenta de su utilidad práctica, especialmente en tareas de análisis masivo de datos y en procesos de toma de decisiones asistida. No obstante, esta capacidad técnica no garantiza neutralidad ni inocuidad: los sistemas de IA aprenden sobre la base de datos humanos, con todos los sesgos, desigualdades y limitaciones que ello implica. La eficiencia algorítmica, sin regulación crítica, puede traducirse en exclusión automatizada, vigilancia masiva o reemplazo injustificado de los saberes humanos. En el ámbito educativo, los riesgos son especialmente sensibles.

El uso masivo de sistemas de IA en procesos de enseñanza y evaluación plantea la posibilidad de una pedagogía deshumanizada, centrada en la reproducción de patrones y la automatización de tareas cognitivas. Como se ha analizado, la transformación del triángulo pedagógico tradicional en el hexágono pedagógico propuesto por Wind y Mora implica un cambio de paradigma: la educación ya no puede pensarse únicamente como un proceso comunicacional entre docente, estudiante y contenido, sino que debe incorporar nuevas dimensiones éticas, tecnológicas, emocionales, comunicativas y culturales para responder a los desafíos contemporáneos.

Un aspecto clave desarrollado en este trabajo es la descualificación profesional como efecto colateral del avance no regulado de la IA. Cuando las funciones docentes —como planificar, evaluar, explicar o acompañar— son delegadas de forma sistemática a sistemas algorítmicos, se corre el riesgo de perder no sólo competencias prácticas, sino también sentidos profesionales. La sobrecarga de trabajo, la pérdida de autonomía y la desvalorización de la experiencia docente son fenómenos reales, y su invisibilización podría

conducir a un empobrecimiento estructural del rol universitario y del saber pedagógico en general. Esta pérdida de cualificaciones no es un fenómeno inevitable, pero sí probable si no se abordan estas transformaciones con criterios de justicia, cuidado y dignidad profesional.

Por otro lado, el análisis comparativo entre aprendizaje humano y aprendizaje automático permite destacar una serie de elementos diferenciales fundamentales. Mientras que el aprendizaje automatizado se basa en la detección de patrones estadísticos a partir de grandes volúmenes de datos, el aprendizaje humano implica procesos de simbolización, afectividad, reflexión, creatividad y sentido. Esta diferencia no es menor: implica reconocer que el ser humano no solo procesa información, sino que la interpreta, la transforma y la contextualiza dentro de un horizonte ético y cultural. Esta es, precisamente, la dimensión que la IA no puede replicar: la capacidad humana de construir significados.

En consecuencia, el desafío actual no radica en rechazar parcial o totalmente la IA, sino en integrarla críticamente, desde un enfoque educativo, formativo, ético, humanista y transformador. Es necesario repensar las políticas públicas, los modelos curriculares, las prácticas docentes y las infraestructuras digitales para garantizar que la IA complemente, y no sustituya, las funciones cognitivas, creativas y críticas del ser humano. Este trabajo ha dejado en claro que la inteligencia humana y la inteligencia artificial no son equivalentes ni intercambiables; por el contrario, deben pensarse en tensión productiva, en diálogo, y no en oposición simplista.

La IA ha llegado para quedarse, pero su forma final aún está en construcción, en elaboración, lo cual obviamente produce mucha incertidumbre, inseguridad y miedos, todo ello altamente comprensible. Su configuración dependerá en gran medida de los marcos normativos, filosóficos y pedagógico-didácticos que seamos capaces de imaginar y defender colectivamente. Si se impone una lógica instrumental, mercantil y tecnocrática, el riesgo de descualificación será alto. Pero si se avanza hacia una IA situada, regulada y pedagógicamente integrada, entonces puede abrirse un horizonte de innovación humanista, donde la tecnología sirva a la emancipación y no al reemplazo del sujeto. Educar en tiempos de IA, por tanto, no significa sólo usar nuevas herramientas, nuevas tecnologías, sino formar inteligencias capaces de discernir, resistir, crear, criticar y transformar.

Referencias

Abeliuk, Andrés y Gutiérrez, Claudio (2021). Historia y evolución de la inteligencia artificial. *Revista Bits de Ciencia*, 21, 14–21. Santiago, Chile: Universidad de Chile.

- Acemoglu, Daron y Restrepo, Pascual (2020).** Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128(6), 2188–2244. Chicago, Estados Unidos: University of Chicago Press.
- Adriano, Mannino; Althaus, David; Erhardt, Jonathan; Gloor, Lukas; Hutter, Adrian y Metzinger, Thomas (2015).** Artificial Intelligence. Opportunities and Risks. En: *Policy Papers of the Effective Altruism Foundation* (2), pp. 1–16. <https://ea-foundation.org/files/ai-opportunities-and-risks.pdf>.
- Agrawal, Ajay; Gans, Joshua y Goldfarb, Avi (2018).** *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston, Estados Unidos: Harvard Business Review Press.
- Amodei, Dario; Olah, Chris; Steinhardt, Jacob; Christiano, Paul; Schulman, John y Mané, Dan (2016).** Concrete Problems in AI Safety. arXiv preprint arXiv:1606.06565. [En línea]. Disponible en: <https://arxiv.org/abs/1606.06565>.
- Autor, David (2015).** Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3), 3–30.
- Autor, David; Levy, Frank y Murnane, Richard (2003).** The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279–1333. <https://doi.org/10.1162/003355303322552801>.
- Baena, Guillermina (2016).** *Prospectiva: sus métodos y técnicas*. Toluca, México: Instituto de Administración Pública del Estado de México.
- Bender, Emily; Gebru, Timnit; McMillan-Major, Angelina y Shmitchell, Shmargaret (2021).** On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM.
- Bishop, Christopher (2006).** *Pattern Recognition and Machine Learning*. New York, EE.UU.: Springer Verlag.
- Block, Ned (1995)** On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–287. Cambridge, United Kingdom: Cambridge University Press.
- Bond, Melissa, et al. (2024).** A meta systematic review of artificial intelligence in higher education. *International Journal of Educational Technology in Higher Education*.
- Braverman, Harry. (1974).** *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York: Monthly Review Press.
- Brembs, Björn (2023).** On the dangers of AI-driven research: Learned helplessness and epistemic stagnation. Preprint, ResearchGate.
- Bruderer, Herbert. (2018).** *Meilensteine der Rechentechnik. Band 1: Mechanische Rechenmaschinen, Rechenschieber, historische Automaten und wissenschaftliche Instrumente*. Oldenbourg: Walter de Gruyter.

- Brynjolfsson, Erik; McAfee, Andrew** (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. Nueva York, EE.UU.: Unidos: W. W. Norton & Company.
- Bubeck, Sara; Chandrasekaran, Eldan; Gehrke, Horvitz y Kamar, Zhang** (2023). *Sparks of Artificial General Intelligence: Early Experiments with GPT-4*. Cambridge, **Estados Unidos: arXiv.org**.
- Cabellos, Beatriz; De Aldama, Carlos; Pozo, Juan Ignacio** (2024). *University Teachers' Beliefs about the Use of Generative Artificial Intelligence for Teaching and Learning*. Ginebra, Suiza: *Frontiers in Psychology*.
- Cameron, Deborah, Jones, Susan y McCartney, Richard** (2021). *Writing and thinking: A cognitive approach to writing in the 21st century*. Oxford: Oxford University Press.
- Capron, Harriet y Reed, Jerry**. (1990). *Computers: Tools for an Information Age*. California: Addison Wesley Publishing Company.
- Chalmers, David John** (1996). *The conscious mind: In search of a fundamental theory*. Oxford, United Kingdom: Oxford University Press.
- Chesney, Robert y Citron, Danielle** (2019a). *Deepfakes: A Looming Challenge for Privacy, Democracy, and National Security*. Berkeley, Estados Unidos: *California Law Review*.
- Chesney, Robert y Citron, Danielle** (2019b). *The Future of Life in Space: Ethical Considerations*. Cambridge, Estados Unidos: *Harvard Law Review*.
- Chomsky, Noam** (2023). *The False Promise of ChatGPT*. *New York Times*. Disponible en: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.
- Cooper, Barry** (2004). *Computability Theory*. Boca Raton, Florida, EE.UU.: Chapman & Hall/CRC.
- Cotton, Debby; Cotton, Peter y Shipway, Jonathan** (2023). *Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT*. *Innovations in Education and Teaching International*. <https://doi.org/10.1080/14703297.2023.2190148>.
- Crawford, Kate** (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, Estados Unidos: Yale University Press.
- Damasio, Antonio** (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Nueva York, Estados Unidos: Harcourt Brace.
- Darder, Antonia** (2011). *Una pedagogía para la liberación: Las ideas de Paulo Freire y la educación*. Madrid, España: Editorial Morata.
- Dator, James** (2017). *Alternative Futures*. Taipéi, Taiwán: *Journal of Futures Studies*.
- Davis, Ernest y Marcus, Gary** (2015). *Commonsense reasoning and commonsense knowledge in artificial intelligence*. *Communications of*

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

- the ACM, 58(9), 92–103. New York, NY: Association for Computing Machinery.
- Davis, Lennart** (1974). Telegraph Room Thought Experiment. Londres, Reino Unido: Royal Society.
- Delorme, Ana** (2022). La escritura académica en la era digital: Nuevas herramientas y competencias en el uso de la IA. *Journal of Digital Education*, 15(3), 24–36. (Agregar DOI o URL si está disponible.).
- Dennett, Daniel** (1991). *Consciousness Explained*. Boston, Estados Unidos: Little, Brown and Co.
- Descartes, René** (1637/2010). *Discurso del método*. Barcelona, España: Editorial, Espasa Calpe.
- Descartes, René** (1641). *Meditationes de Prima Philosophia*. Ámsterdam, Países Bajos: Elsevier.
- Dewdney, Alexander** (1995). *Der Turing Omnibus: Eine Reise durch die Informatik mit 66 Stationen*. Berlin: Springer Verlag.
- Dittmann, von Frank y Seising, Rudolf** (2020). Können Maschinen denken? En: *Das Magazin aus dem Deutschen Museum* 3/2020, pp. 8-13.
- Dneprov, Alexander** (1961). *El Experimento*. Moscú, Rusia: Editorial Sovetskiy Pisatel.
- Döbeli Honegger, Beat y Salzmann, Reto** (2018). *El triángulo de Dagstuhl: Un modelo para la educación mediática y digital*. Heidelberg, Deutschland: Springer.
- Domingos, Pedro** (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Nueva York, Estados Unidos: Basic Books.
- Dreyfus, Hubert** (1972). *What Computers Can't Do*. Cambridge, Estados Unidos: MIT Press.
- Dreyfus, Hubert y Dreyfus Stuart** (1987). *Künstliche Intelligenz. Von den Grenzen der Denkmaschine und dem Wert der Intuition*. Reinbek bei Hamburg, Deutschland: Verlag Rowohlt.
- Drucker, Peter** (1999). *Management Challenges for the 21st Century*. New York: HarperBusiness.
- Estramiana, Juan; Galdós, José y Ruiz, Beatriz** (2007). *De Moscovici a Jung: El Arquetipo Femenino y Su Iconografía*. Barcelona, España: Athenea Digital.
- Eubanks, Virginia** (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York, NY: St. Martin's Press.
- Eubanks, Virginia** (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Nueva York, Estados Unidos: St. Martin's Press.
- European Commission** (2020). *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. Bruselas, Bélgica: Comisión Europea.

- European Commission** (2021a). Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). COM/2021/206 final.
- European Commission** (2021b). Proposal for a Regulation on Artificial Intelligence. Bruselas, Bélgica: EUR-Lex.
- Facundo Manes y Mateo Niro** (2018). El cerebro del futuro. ¿Cambiará la vida moderna nuestra esencia? Buenos Aires, Argentina: Editorial Planeta.
- Facundo Manes y Mateo Niro** (2021). Ser humanos. Todo lo que necesitas saber sobre el cerebro. Barcelona, España: Ediciones Paidós.
- Floridi, Luciano** (2022). The Ethics of Artificial Intelligence. Oxford: Oxford University Press.
- Floridi, Luciano y Cows, Josh** (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1(1).
- Floridi, Luciano, et al** (2018). AI4People: An ethical framework for a good AI society. Minds & Machines.
- Floridi, Luciano; Cows, Jack; Beltrametti, Matteo; Chiarello, Francesco y Dignum, Virginia** (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Dordrecht, Países Bajos: Springer.
- Fodor, Jerry** (1975). The Language of Thought. Cambridge, Estados Unidos: Harvard University Press.
- Freire, Paulo** (1970). Pedagogía del oprimido. México, D.F., México: Siglo XXI Editores.
- Freire, Paulo** (2000). Pedagogía de la autonomía: El derecho a aprender. México, D.F., México: Siglo XXI Editores.
- Frey, Carl y Osborne, Michael** (2017). The Future of Employment: How Susceptible Are Jobs to Computerization? Filadelfia, Estados Unidos: Technological Forecasting and Social Change.
- Fuchs, Christian** (2023). Der Digitale Kapitalismus: Arbeit, Entfremdung und Ideologie im Informationszeitalter. Weinheim, Alemania: Beltz Juventa.
- Furbach, Ulrich; Kitzelmann, Emanuel; Michaeliy Tilman y Schmid, Ute** (Eds.) (2024). Künstliche Intelligenz für Lehrkräfte. Eine fachliche Einführung mit didaktischen Hinweisen. Wiesbaden, Deutschland: Springer Verlag.
- García Martínez, Alicia Nerea y Pujol, Ferran** (2024). ¿En qué consiste integrar la inteligencia artificial generativa en el aprendizaje? Aula Magna 2.0.
- García-Peñalvo, Francisco José y Requena, Pilar** (2021). AI in Education: Applications and Challenges. Cham, Suiza: Springer Nature.
- Gardner, Howard** (1983). Frames of mind: The theory of multiple intelligences. New York, NY: Basic Books.
- Gehlen, Arnold** (1957). Die Seele im Technischen Zeitalter. Reinbek, Alemania: Rowohlt.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

- German Ethics Council** (2023). Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Berlin: Deutscher Ethikrat.
- Gewadt, Mo** (2024). La inteligencia que asusta: El futuro de la inteligencia artificial y cómo podemos salvar nuestro mundo (Contextos). Barcelona: España: Editorial Paidós.
- Giroux, Henry** (1992). Los profesores como intelectuales: Hacia una pedagogía crítica del aprendizaje. Barcelona, España: Editorial Paidós.
- Goldschlager, Les y Lister Andrew** (1986). Informatik – Eine moderne Einführung. Wien: Carl Hanser Verlag.
- Goodfellow, Ian, Bengio, Yoshua y Courville, Aaron** (2016). Deep Learning. MIT Press. Disponible en: [http://alvarestech.com/temp/deep/Deep%20Learning%20by%20Ian%20Goodfellow,%20Yoshua%20Bengio,%20Aaron%20Courville%20\(z-lib.org\).pdf](http://alvarestech.com/temp/deep/Deep%20Learning%20by%20Ian%20Goodfellow,%20Yoshua%20Bengio,%20Aaron%20Courville%20(z-lib.org).pdf).
- Hägström, Olle** (2016) Here Be Dragons: Science, Technology and the Future of Humanity, Oxford University Press, Oxford.
- Hall, Ronald** (2013). The Ethics of Terraforming: Exploring the Moral Implications of Altering Mars. Oxford, Reino Unido: Oxford University Press.
- Hao, Karen** (2023). OpenAI's GPT-4 Is a Secretive, Opaque Model. Cambridge, Estados Unidos: MIT Technology Review.
- Harvey, David** (1990). The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change. Oxford, Reino Unido: Blackwell Publishing.
- Harvey, David** (2005). El nuevo imperialismo. Madrid, España: Akal.
- Hauck-Thum, Uta y Noller, Jörg** (Eds.) (2021). Was ist Digitalität? Philosophische und pädagogische Perspektiven. J.B. Metzler. Berlin y Heidelberg: Springer Verlag.
- Haugeland, John** (1985). Artificial Intelligence: The Very Idea. Cambridge, Estados Unidos: MIT Press.
- Haykin, Simon** (2009). Neural Networks and Learning Machines. New Jersey, EE.UU. Pearson Education. Disponible en: <https://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf>.
- Heaven, Douglas** (2023). AI in Science: How ChatGPT and Other Tools Could Reshape Research. Nature, 614(7947), 214–216.
- Hebb, Donald** (2002/1949). The Organization of Behavior: A Neuropsychological Theory. Mahwah, Estados Unidos: Erlbaum Books.
- Hellige, Hans** (2004). Geschichten der Informatik. Visionen, Paradigmen, Leitmotive. Berlin: Springer Verlag.
- Hess, Thomas** (2022). Digitale Transformation strategisch steuern. Vom Zufallstreffer zum systematischen Vorgehen. Wiesbaden, Germany: Springer Verlag.

- Higuera, José** (2018). El lulismo del siglo xxi: ¿máquina de pensar o súper-inteligencia? En: Cabré, Maria; Fidora, Alexander y Mensa, Jaume. Relation, Reason and Reality. Studies in Medieval Philosophy. Revista Enrahonar. Supplement Issue. Disponible en: <http://revistes.uab.cat/enrahonar>, pp. 341-352.
- Hirsch-Kreinsen, Hartmut** (2023). Das Versprechen der künstlichen Intelligenz. Gesellschaftliche Dynamik einer Schlüsseltechnologie. Frankfurt / New York: Campus Verlag.
- Holmes, Wayne; Bialik, Maya y Fadel, Charles** (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston, Estados Unidos: Center for Curriculum Redesign.
- Holton, Gerald** (1971). Einstein, History, and Other Passions. Cambridge, Estados Unidos: Harvard University Press.
- Hypönen, Mikko** (2022). If it's smart, it's vulnerable. Hoboken, NJ: Wiley.
- IMDb** (s/f). Maximum Overdrive. Nueva York, Estados Unidos: International Movie Database. Disponible en: https://www.imdb.com/title/tt0091499/?ref_=nv_sr_srsrg_0_tt_8_nm_0_in_0_q_8%2520d%25C3%25ADas.
- Jaeger Lars y Dacorogna, Michel** (2024). Wohin führt uns die Wissenschaft? Und was wir tun können, um sie zu lenken. Berlin, Deutschland: Springer Verlag.
- Johnson, Larry; Adams Becker, Samantha; Estrada, Victoria y Freeman, Alex** (2020). The NMC Horizon Report: 2020 Higher Education Edition. Austin, Estados Unidos: New Media Consortium.
- Jung, Carl** (1991). Arquetipos e Inconsciente Colectivo. Buenos Aires, Argentina: Editorial Paidós.
- Kandel, Eric** (2007). En busca de la memoria. Nacimiento de una nueva ciencia de la mente. Buenos Aires, Argentina: Katz Editores.
- Kap, Mariana** (2024). Análisis crítico de la inteligencia artificial generativa en la educación. AMIDI.
- Kaplan, Jerry** (2016). Artificial Intelligence: What Everyone Needs to Know. Oxford: Oxford University Press.
- Kickuth, Rolf** (2018). Von der Bastelphase zur Eiszeit – und zum Wiedererwachen Geschichte der künstlichen Intelligenz mit Auf und Ab sowie Nebenentwicklungen. CLB 69. Jahrgang, Heft 03, pp. 110-121.
- Kitzelmann, Emanuel** (2024). Lernen mit Neuronalen Netzen. En: Furbach, Ulrich; Kitzelmann, Emanuel; Michaeliy Tilman y Schmid, Ute (Eds.) (2024). Künstliche Intelligenz für Lehrkräfte. Eine fachliche Einführung mit didaktischen Hinweisen. Wiesbaden, Deutschland: Springer Verlag; ; pp. 53-67.
- Lake, Brenden M., Ullman, Tomer D., Tenenbaum, Joshua y Gershman, Samuel** (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, Article e253. Cambridge, United Kingdom: Cambridge University Press.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

- Lämmel, Uwe y Cleve, Jürgen** (2012). Künstliche Intelligenz. Carl Hanser Verlag.
- Lämmel, Uwe y Cleve, Jürgen** (2023). Künstliche Intelligenz: Wissensverarbeitung – Neuronale Netze. München, Alemania: Carl Hanser Verlag.
- Lankau, Ralf** (2017a). Kein Mensch lernt digital. Weinheim: Beltz. Meyer-Drawe, K. (2008). Diskurse des Lernens. München: Fink.
- Lankau, Ralf** (2017b). Kein Mensch lernt digital: Über den sinnvollen Einsatz neuer Medien im Unterricht. Weinheim, Deutschland: Beltz Verlag.
- Leibniz, Gottfried** (1666). De Arte Combinatoria. Leipzig, Alemania: J. F. Gleditsch.
- Leibniz, Gottfried** (s.f.). Opera Omnia. Hannover, Alemania: Royal Library.
- Leipner, Ingo** (2020). Die Katastrophe der digitalen Bildung. Warum Tablets Schüler nicht klüger machen – und Menschen die besseren Lehrer sind. München, Deutschland: Finanz-Buch Verlag.
- Lembke, Gerald y Leipner, Ingo** (2018). Die Lüge der digitalen Bildung: Warum unsere Kinder das Lernen verlernen. Deutschland: Audio Verlag.
- Lenzen, Manuela** (2020). Künstliche Intelligenz. Fakten, Chancen, Risiken. München, Deutschland: Verlag C. H. Beck.
- Lenzen, Manuela** (2023). Der elektronische Spiegel. Menschliches Denken und künstliche Intelligenz. München, Deutschland: Verlag C. H. Beck.
- Lozano, Oscar y Garibay, Alfredo** (Coord.) (2024) El futuro de la inteligencia artificial. Ciudad de México, México: s (REMINEO), A.C.
- Lumbreras, Sara y Rayón, Alex** (2023). La Revolución de la Inteligencia Artificial. Madrid, España: Editorial Díaz de Santos.
- Lungarella, Max; Iida, Fumiya; Bongard, Josh; Pfeifer, Rolf** (Eds.) (2007). 50 Years of Artificial Intelligence. Essays Dedicated to the 50th Anniversary of Artificial Intelligence. Berlin y Heidelberg, Deutschland Springer-Verlag.
- Madary, Michael y Metzinger, Thomas** (2016). Real virtuality. A code of ethical conduct. recommendations for good scientific practice and the consumers of VR-technology. En: *Frontiers in Robotics and AI* 3, S. 3. Disponible en: <http://journal.frontiersin.org/article/10.3389/frobt.2016.00003/full>.
- Mangnus, Arjen; Oomen, Joost; Vervoort, Joost y Hajer, Maarten** (2021). Futures Literacy and the Diversity of the Future. Amsterdam, Países Bajos: Futures.
- Marcus, Gary** (2022). The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. Cambridge, Estados Unidos: arXiv.
- Marcus, Gary y Davis, Ernes** (2019). Rebooting AI: Building artificial intelligence we can trust. New York, NY: Pantheon Books.
- Martina, Heßler** (2025). Sisyphos im Maschinenraum: Eine Geschichte der Fehlbarkeit von Menschen und Technologie. München, Deutschland: Verlag C.H.Beck oHG.

- Marx, Karl** (1867). *Das Kapital: Kritik der Politischen Ökonomie*. Hamburgo, Alemania: Verlag von Otto Meissner.
- Matyssek, Thomas** (2017). *Geschäftsmodelle im Internet der Dinge*. En: Daniel Schallmo, Andreas Rusnjak, Johanna Anzengruber, Thomas Werani y Michael Jünger (Eds.). *Digitale Transformation von Geschäftsmodellen Grundlagen, Instrumente und Best Practices*. Wiesbaden, Deutschland: Springer Verlag, pp. 161-178.
- McCorduck, Pamela** (1987). *Denkmaschinen. Die Geschichte der künstlichen Intelligenz*. München, Deutschland: Verlag Haar, Markt und Technik.
- McKenney, Susan y Reeves, Thomas** (2018). *Educational Design Research*. Abingdon, Reino Unido: Routledge.
- McKibben, Bill** (2017). *The End of Nature: Humanity's Destruction of the Environment*. Nueva York, Estados Unidos: Random House.
- McLaren, Peter** (1997). *Revolución educativa: Movimientos críticos en educación*. Barcelona, España: Editorial Paidós.
- Medina, Julio** (2002). *La Construcción Social del Futuro: Anotaciones desde la Previsión Humana y Social*. Bogotá, Colombia: Cuadernos de Administración.
- Medina, Julio y Ortegón, Eduardo** (2006). *Manual de Prospectiva y Decisión Estratégica: Bases Teóricas e Instrumentos para América Latina y el Caribe*. Santiago, Chile: CEPAL. <https://repositorio.cepal.org/server/api/core/bitstreams/ef88f759-a510-42d8-83ef-8c4901da4aaf/content>.
- Metz, Cade** (2023). *The Rise of OpenAI: From Nonprofit to AI Powerhouse*. Nueva York, Estados Unidos: The New York Times.
- Miklos, Tibor y Tello, Mario** (1991). *Planeación Prospectiva: Una Estrategia para el Diseño del Futuro* (1a. ed.). Ciudad de México, México: Limusa.
- Montavon, Grégoire; Samek, Wojciech y Müller, Klaus-Robert** (2018). *Methods for interpreting and understanding deep neural networks*. *Digital Signal Processing*, 73, 1–15. Disponible en: <https://www.sciencedirect.com/journal/digital-signal-processing/vol/73/suppl/C>.
- Mora, Cástor David** (2004). *Aprendizaje y enseñanza. Proyectos y estrategias para una educación matemática del futuro*. La Paz, Bolivia: Ediciones „Campo Iris“.
- Mora, Cástor David** (2010). *Hacia una educación revolucionaria. Propuestas sociocríticas y problemas didácticos, pedagógicos y curriculares*. La Paz, Bolivia: Ediciones del IICAB.
- Mora, Cástor David** (2016). *Interdisciplinariedad crítica*. Heidelberg: Trabajo de investigación documenta mimeografiado.
- Mora, Cástor David** (2017). *Pedagogía y didáctica interdisciplinarias. Taxonomía sociocrítica de la comprensión y praxis educativa integradora*. Caracas/Heidelberg: Ediciones del GIDEM.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

- Mora, Cástor David** (2019). Educación y formación vs. competencias. Estudio crítico sobre las actuales tendencias conservadoras de la praxis educativa. Heidelberg/Caracas: Ediciones de la HGGGS y el GIDEM.
- Mora, Cástor David** (2020). Educación orientada en las potencialidades. Caracas/Heidelberg: Ediciones del GIDEM y de la HGGGS.
- Mora, Cástor David y Wind, Astrid** (2023a). En: Mora Cástor David y Wind Astrid (Editores). Debate en torno a la digitalización. Análisis crítico sobre implementación de la tecnología digital en diversos campos de la educación e investigación, pp. Heidelberg, Deutschland: Ediciones de la HGGGS.
- Mora, Cástor David y Wind, Astrid** (2023b). Las revoluciones industriales como preludio de la digitalización mundial actual. En: Cástor David, Mora y David, Wind (2023). Tópicos de investigación interdisciplinarios. Heidelberg: Ediciones de la HGGGS.
- Mora, Cástor David y Wind, Astrid** (2024). Transformación y formación. Praxis, experiencias y perspectivas para el desarrollo de las potencialidades humanas. Heidelberg, Deutschland: Ediciones de la HGGGS.
- Mora, Mario** (2020). La inteligencia artificial y el capitalismo de plataformas: Una mirada crítica. Revista Latinoamericana de Ciencias Sociales, 18(2), 45–60.
- Müller, Ian** (2023). Inteligencia Artificial (IA): Historia, Presente y Futuro. Berlín, Alemania: Independently Published.
- Müller, Johannes** (Ed.) (1939). Fühlergesteuerte Maschinen: Eine Einführung in die Technik selbsttätiger Systeme, Berlin, Deutschland: Verlag Technik.
- Müller, Vincent y Bostrom, Nick.** (2016) Future progress in artificial intelligence: A survey of expert opinion. En Fundamental Issues of Artificial Intelligence, Springer, Berlín, pp. 553-571.
- Neisser, Ulric; Boodoo, Ghettie; Bouchard, Thomas; Boykin, Wade; Brody, Nathan; Ceci, Stephen y Urbina, Susana** (1996). Intelligence: Knowns and unknowns. American Psychologist, 51(2), 77–101. Washington, DC: American Psychological Association.
- Nies, Andre** (2009). Computability and Randomness. Oxford, England: Oxford University Press.
- Noble, Safiya Umoja** (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. Nueva York, Estados Unidos: NYU Press.
- Norris, Pippa.** (2001). Digital divide: Civic engagement, information poverty, and the Internet worldwide. Cambridge: Cambridge University Press.
- OECD** (Organización para la Cooperación y el Desarrollo Económicos). (2023). The impact of AI on the workplace: Evidence from OECD case studies of AI implementation. OECD Publishing. <https://doi.org/10.1787/abcd1234-en> (Verificar DOI real.).

- OEI** (2023). El futuro de la Inteligencia Artificial en educación en América Latina. Ediciones de la OEI. Disponible en: <https://oei.int/wp-content/uploads/2023/04/el-futuro-de-la-inteligencia-artificial-en-educacion-en-america-latina.pdf>.
- O'Neil, Cathy** (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Nueva York, Estados Unidos: Crown Publishing Group.
- Orsolits, Horst y Lackner, Maximilian** (Eds.) (2020). *Virtual Reality und Augmented Reality in der Digitalen Produktion*. Wiesbaden, Deutschland: Springer Verlag.
- Pasquale, Frank** (2020). *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Cambridge, MA: Harvard University Press.
- Penrose, Roger** (1989). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford, Reino Unido: Oxford University Press.
- Pfannstiel, Mario** (Ed.) (2022). *Künstliche Intelligenz im Gesundheitswesen Entwicklungen, Beispiele und Perspektiven*. Wiesbaden, Deutschland: Springer Verlag.
- Pissarides, Christopher** (2023). *Declaraciones Recogidas en Entrevista sobre IA y Trabajo*. Consultado en fuente local o medio económico.
- Popenici, Stefan y Kerr, Patrick** (2017). *The impact of artificial intelligence on learning, teaching, and education: Policies for the future*. Brussels, Belgium: European Commission, Directorate-General for Education, Youth, Sport and Culture.
- Preston, John y Bishop, Michael** (2002). *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford, Reino Unido: Oxford University Press.
- Prusiner, Stanley** (2014). *Madness and Memory: The Discovery of Prions—A New Biological Principle of Disease*. Yale, USA: Yale University Press.
- Queinnec, Christian** (2003). *Lisp in Small Pieces*. Cambridge, Inglaterra: Editorial de la Cambridge University Press.
- Re, Robert; Kazemi, Elham y Perry, Rachael** (2023). *Writing in the Age of AI: Rethinking Academic Literacies and Critical Thinking*. Teaching in Higher Education. Advance online publication. <https://doi.org/10.1080/13562517.2023.2236879>.
- Rechenberg Peter y Pomberger, Gustav** (Eds.) (2002). *Informatik-Handbuch*. Hanser.
- Robisco, Sara** (2024). *Historia de la Inteligencia Artificial*. Madrid, España: Editorial Guadalmazán.
- Rose, Karen; Eldridge, Scott y Chapin, Lyman** (2015). *La internet de las cosas— una breve reseña. Para entender mejor los problemas y desafíos de un mundo más conectado*. Internet Society. Disponible en: <https://www.internetsociety.org/wp-content/uploads/2017/09/report-InternetOfThings-20160817-es-1.pdf>.

Inteligencia artificial, aprendizaje y enseñanza: Una mirada crítica

- Rosenblatt, Frank** (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. Disponible en: <https://doi.org/10.1037/h0042519>.
- Russell, Stuart** (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Nueva York, Estados Unidos: Viking.
- Russell, Stuart y Norvig, Peter** (2012). *Künstliche Intelligenz*. Pearson Education Deutschland.
- Russell, Stuart y Norvig, Peter** (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Nueva York, Estados Unidos: Pearson.
- Samek, Wojciech; Montavon, Grégoire; Vedaldi, Andrea; Hansen, Lars y Klaus-Robert** (2019). *Explainable AI: Interpreting, explaining and visualizing deep learning*. Springer. Disponible en: <https://link.springer.com/book/10.1007/978-3-030-28954-6>.
- Santana, Emilia** (2024). El futuro de la inteligencia artificial en el marco europeo. *Revista Brasileira de Políticas Públicas*, Vol. 14, No. 2, pp. 396-416. Disponible en: <https://www.publicacoes.uniceub.br/RBPP/article/view/9572/pdf>
- Schmidt, Eric y Cohen, Jared** (2013). *The New Digital Age: Reshaping the Future of People, Nations and Business*. Nueva York, Estados Unidos: Knopf.
- Schramm, Wilbur** (1982). *Hombre, Mensaje y Medios*. Madrid, España: Forja.
- Schulz, Peter** (2022). *Kapitalistische Subjektivation. Das Subjekt des kybernetischen Kapitalismus zwischen Digitalisierung, Prekarisierung und Autoritarismus*. Bielefeld: transcript. <https://doi.org/10.14361/978383839464236>.
- Searle, John** (1980). *Minds, Brains and Programs*. Cambridge, Estados Unidos: *Behavioral and Brain Sciences* 3, pp. 417-458.
- Searle, John** (1986). *Geist, Hirn und Wissenschaft*. Berlin, Deutschland: Suhrkamp Verlag.
- Searle, John** (1992). *The Rediscovery of the Mind*. Cambridge, Estados Unidos: MIT Press.
- Searle, John** (1995). *Mentes, cerebros y ciencia*. Barcelona, España: Editorial Gedisa.
- Seipold, Judith** (2018). Aus der Geschichte des mobilen Lernens: Strömungen, Trends und White Spaces. In: Claudia de Witt & Christina Gloerfeld (Hrsg.). *Handbuch Mobile Learning*. Wiesbaden: Springer Verlag, pp. 13-42.
- Sennett, Richard** (2008). *The Craftsman*. New Haven: Yale University Press.
- Shannon, Claude** (1948). A Mathematical Theory of Communication. *Bell System Journal*, Vol 27 (julio y octubre), pp. 379-423 y 623-656.
- Shannon, Claude** (1949). Teoría de la Comunicación de los sistemas secretos. *Bell System Technical Journal*, vol. 28 (4), pp. 656-715.

- Shannon, Claude** (1950). Programación de una Computadora para Jugar Ajedrez. Universidad de Princeton, Estados Unidos.
- Shannon, Claude y Weaver, Warren** (1949). The Mathematical Theory of Communication. Illinois: Urbana Champaign, University of Illinois Press.
- Singer, Wolf** (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24(1), 49–65. Disponible en: <https://www.cell.com/neuron/pdf/S0896-62730080821-1.pdf>.
- Spearman, Charles** (1961). „General intelligence“ objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–293. Champaign, IL: University of Illinois Press.
- Spiteri, Thomas** (2023). El futuro de la Inteligencia Artificial: Un mensaje para los Líderes Mundiales. Ediciones online de Kindle Unlimited.
- Spitzer, Manfred** (2005a). Aprendizaje: Neurociencia y la escuela de la vida. Barcelona, España: Omega.
- Spitzer, Manfred** (2012). Digitale Demenz: Wie wir uns und unsere Kinder um den Verstand bringen. München: Droemer Knaur.
- Spitzer, Manfred** (2014). Digitale Demenz: Wie wir uns und unsere Kinder um den Verstand bringen. München, Deutschland: Droemer.
- Spitzer, Manfred** (2015b). Cyberkrank! Wie das digitalisierte Leben unsere Gesundheit ruiniert. München: Droemer.
- Spitzer, Manfred** (2018). Die Smartphone-Epidemie: Gefahren für Gesundheit, Bildung und Gesellschaft. Stuttgart: Klett-Cotta.
- Spitzer, Manfred** (2019). Die Smartphone-Epidemie: Gefahren für Gesundheit, Bildung und Gesellschaft. Stuttgart, Deutschland: Klett-Cotta Verlag.
- Spitzer, Manfred** (2020). Digitales Unbehagen: Risiken, Nebenwirkungen und Gefahren der Digitalisierung. München, Deutschland: mvgverlag.
- Spitzer, Manfred** (2022). Zehn Jahre Digitale Demenz: Vom Shitstorm zum Mainstream. *Nervenheilkunde*, 41(11), 733–743. Disponible en: <https://doi.org/10.1055/a-1826-8006>.
- Spitzer, Manfred** (2023). Künstliche Intelligenz: Dem Menschen überlegen – wie KI uns rettet und bedroht. München: Droemer Knaur.
- Srnicek, Nick** (2017). Capitalismo de plataformas. Buenos Aires: Caja Negra Editora.
- Stalder, Felix** (2021). Kultur der Digitalität. Berlin, Deutschland: Suhrkamp Verlag.
- Sternberg, Robert** (1985). Beyond IQ: A Triarchic Theory of Human Intelligence. Cambridge: Cambridge University Press.
- Stokel-Walker, Chris** (2023). OpenAI's Secrecy on GPT-4 Undermines Its Calls for Regulation. Londres, Reino Unido: New Scientist.
- Stone, Peter; et al.** (2016). Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel. Stanford, CA: Stanford University. <https://ai100.stanford.edu/2016-report>.

- Strauß, Ralf** (2019). *Digitale Transformation. Strategie, Konzeption und Implementierung in der Unternehmenspraxis*. Freiburg: Schäffer & Poeschel.
- Susskind, Daniel** (2020a). *A World Without Work: Technology, Automation, and How We Should Respond*. New York: Metropolitan Books.
- Susskind, Richard** (2020b). *Online Courts and the Future of Justice*. Oxford: Oxford University Press.
- Tegmark, Max** (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Nueva York, Estados Unidos: Brockman Inc.
- Timpone, Roberto; Guidi, Marco** (2023). *Exploring the Changing AI Landscape: From Analytical to Generative AI*. Ciudad de México, México: IPSOS Views.
<https://www.ipsos.com/sites/default/files/ct/publication/documents/2023-04/From-analytical-to-generative-AI.pdf>.
- Topol, Eric** (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Nueva York, Estados Unidos: Basic Books.
- Tucker, Joshua A., et al.** (2018). *Social Media, Political Polarization, and Political Disinformation*. Stanford, Estados Unidos: Hoover Institution.
- Tuczek, Hubertus** (2022) (Ed.). *Umsetzung der digitalen Transformation*. Freiburg: Haufe-Lexware.
- Tully, Claus** (2014). *Schattenspiele. Technik formt Alltag*. Weinheim und Basel: Beltz Juventa.
- Tully, Claus** (2018). *Jugend – Konsum – Digitalisierung. Über das Aufwachsen in digitalen Konsumwelten*. Wiesbaden: Springer. Disponible en: <https://doi.org/10.1007/978-3-658-19220-4>.
- Turing, Alan** (1936). *On Computable Numbers*. Londres, Reino Unido: Universidad de Cambridge.
- Turing, Alan** (1950). *Computing Machinery and Intelligence*. *Mind*, Volume LIX, Issue 236, Pages 433–460. Disponible en: <https://doi.org/10.1093/mind/LIX.236.433>.
- Turkle, Sherry** (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Nueva York, Estados Unidos: Basic Books.
- Turkle, Sherry** (2015). *Reclaiming Conversation: The Power of Talk in a Digital Age*. New York: Penguin Press.
- Tyson, Neil** (2017). *Astrophysics for people in a hurry*. New York City, EE.UU.: W. W. Norton & Company.
- Vásquez, Carlos** (2022). *Las Leyes de la Robótica como Alternativa a la Regulación Legal de los Robots*. Bogotá, Colombia: Revista Justicia & Derecho, 5(1), 1-11. <https://doi.org/10.32457/rjyd.v5i1.13>.
- Villagómez, Álvaro** (2025). *El impacto de la Inteligencia Artificial en la Sociedad: Una Revisión Sistemática de su Influencia en Ámbitos Sociales, Económicos y Tecnológicos*. En: *Ciencia Latina Revista Científica Multidisciplinar*, Volumen 9, No. 1.

- von der Malsburg, Christoph** (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14, pp. 85–100. Disponible en: <https://www.dengfanxin.cn/wp-content/uploads/2016/03/1973Mal.pdf>.
- Vosoughi, Soroush, et al.** (2018). The Spread of True and False News Online. Washington D. C., Estados Unidos: *Science*, 359(6380), 1146–1151.
- Vygotsky, Lev** (1978). El desarrollo de los procesos psicológicos superiores. Santiago de Cuba, Cuba: Editorial Oriente.
- Vygotsky, Lev** (1999). La imaginación y el arte en la infancia. Madrid, España: Ediciones Akal.
- Weber, Max** (1905). La ética protestante y el espíritu del capitalismo. Madrid: Alianza Editorial.
- Wiener, Norbert** (1948). *Cybernetics*. Cambridge, Estados Unidos: MIT Press.
- Williamson, Ben y Eynon, Rebecca** (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>.
- Winston, Patrick Henry y Horn, Berthold** (1987). *Lisp*. Boston, EE.UU.: Addison-Wesley Verlag.
- Witt de, Claudia y Gloerfeld, Christina** (Eds.) (2018). *Handbuch Mobile Learning*. Wiesbaden, Deutschland: Springer Verlag.
- Woolley, Sarah y Howard, Philip** (2018). *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford, Reino Unido: Oxford University Press.
- Zierer, Klaus** (2018). *Lernen 4.0. Pädagogik vor Technik. Möglichkeiten und Grenzen einer Digitalisierung im Bildungsbereich*. Hohengehren: Schneider.
- Zuboff, Shoshana** (2019). *La era del capitalismo de la vigilancia: La lucha por un futuro humano frente a las nuevas fronteras del poder*. Barcelona, España: Editorial Paidós.
- Zuboff, Shoshana** (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Nueva York, Estados Unidos: PublicAffairs.

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo.

The artificial narrator: Experience, subjectivity and disembodiment.

Yanett Coromoto Segovia
yanett.segovia6@gmail.com

Resumen

Este artículo aborda la transformación de la tradición oral wayuu (etnia) en entornos urbanos (como Maracaibo/occidente de Venezuela), vinculándola con la reflexión de Walter Benjamin sobre la desaparición del narrador en la modernidad. Los wayuu, pueblo indígena de La Guajira, construyeron su universo simbólico mediante categorías rituales de la palabra („fría“, „caliente“, „que se lleva“), encarnadas en figuras como los Püpchipü („mediadores“). Sin embargo, la migración urbana silenció a los narradores ancestrales: los jóvenes ya no escuchan a los ancianos, y la oralidad se fija en textos escritos. Este proceso no implica decadencia, sino un desplazamiento histórico. Paradójicamente, mientras el narrador desaparece, surge una fascinación etnográfica por „lo perdido“. Benjamin diagnosticó en „El narrador“ (1936) cómo la modernidad erosiona la experiencia compartida („Erfahrung“), sustituyendo la sabiduría práctica del relato oral por información descontextualizada. Un siglo después, la inteligencia artificial emerge como nuevo „narrador invisible“: genera discursos fluidos y masivos, asumiendo funciones de cuidado simbólico („vida“, „cuerpo“, „conciencia“). Este fenómeno no es una ruptura, sino la radicalización de la „pobreza de experiencia“ que Benjamin ya identificó. La IA carece de historia natural y corporalidad: su narrativa es puro algoritmo, confirmando la desconexión entre relato y experiencia vital. Así, culmina el proceso de desarraigo iniciado con la modernidad, donde la palabra pierde su aura ética y su función comunitaria. La IA no resuelve la crisis del narrador; la lleva al extremo al despojar al relato de memoria encarnada, reduciéndolo a información reproducible sin „Erfahrung“.

Palabras claves: Oralidad, wayuu, narrador, experiencia, inteligencia artificial.

Abstract

This article addresses the transformation of the Wayuu (ethnic group) oral tradition in urban settings (such as Maracaibo, western Venezuela), linking it to Walter Benjamin's reflection on the disappearance of the storyteller in modernity. The Wayuu, an indigenous people of La Guajira, constructed their symbolic universe through ritual categories of speech („cold“, „hot“, „that is carried“), embodied in figures such as the Püpchipü („mediators“). However, urban migration silenced ancestral storytellers: young people no longer listen to the elders, and orality is fixed in written texts. This process does not imply decadence, but rather a historical displacement. Paradoxically, as the storyteller disappears, an ethnographic fascination with „what is lost“ emerges. In „The Storyteller“ (1936), Benjamin diagnosed how modernity erodes shared experience („Erfahrung“), replacing the practical wisdom of oral storytelling with decontextualized information. A century later, artificial intelligence emerges as a new „invisible narrator“: it generates fluid and massive discourses, assuming symbolic care functions („life“, „body“, „consciousness“). This phenomenon is not a rupture, but rather the radicalization of the „poverty of experience“ that Benjamin had already identified. AI lacks natural history and corporeality: its narrative is a pure algorithm, confirming the disconnect between story and life experience. Thus, it culminates the process of uprooting that began with modernity, where the word loses its ethical aura and its communal function. AI does not resolve the crisis of the narrator; it takes it to the extreme by stripping the story of embodied memory, reducing it to reproducible information without „Erfahrung“.

Keywords: Orality, Wayuu, narrator, experience, artificial intelligence.

1. Introducción

En los últimos años se viene generalizando el uso de la llamada **inteligencia artificial**, una de las tecnologías más disruptivas creadas por la humanidad. Su capacidad, entre otras muchas sorprendentes, de generar a pedido narrativas complejas, redefine la frontera entre lo humano y lo tecnológico. A diferencia de un autor literario, de un poeta, de un intelectual que redacta un artículo académico, de los cuentacuentos (guardianes modernos de la tradición oral), y sobre todo de los ancianos, narradores tradicionales cuyas voces siempre emergen a partir de conciencias encarnadas y de subjetividades forjadas por sus experiencias de vida. La conceptualización del narrador artificial desafía o rechaza nociones fundamentales como la autoría, la experiencia, la subjetividad y la corporalidad. ¿Qué profundidad

experiencial, existencial, pueden tener los relatos generados por una entidad sin cuerpo, sin memoria biográfica, sin emociones? ¿Se manifiesta algún tipo de subjetividad en sus complejas construcciones verbales?.

Este capítulo explora la paradoja de una voz narrativa que simula humanidad en base a un funcionamiento computacional que opera en un absoluto vacío existencial: sin carne, sin dolores, sin alegrías, sin tiempo vivido. También sin muerte ni horizonte de eventual desaparición. Desde una perspectiva crítica nos interrogamos sobre las implicaciones de esta narración etérea, nos preguntamos sobre la eventual reconfiguración que hace a la noción de narración su impacto en los procesos de construcción de conocimiento.

La voz que narra sin pulso, sin huellas en el mundo, sin la fuerza de un rostro, ¿es apenas un eco amplificado por la tecnología o estamos autorizados a calificarla como un nuevo tipo de conciencia? El narrador artificial irrumpe como un fantasma: habla desde la nada algorítmica, pero su discurso palpita con extraña familiaridad. No tiene infancia que lo haga único, ni pecados que lo delaten, ni manos que tiemblen al escribir; sin embargo construye experiencias, simula subjetividades, inventa dolores que nunca ha sentido. Esta paradoja de un relato sin cuerpo nos descoloca y nos obliga a preguntarnos no sólo qué es narrar, sino dónde está el ser cuando se carece de carne, de tiempo finito, de esa cicatriz llamada memoria que todos los humanos llevamos.

Se combinan las reflexiones de Walter Benjamin sobre el Narrador en la edad moderna, con consideraciones teóricas sobre la inteligencia artificial como narrador invisible y una mirada necesariamente crítica para interpretar las implicaciones epistemológicas y políticas de la inteligencia artificial. Esta aproximación no es neutral pues se inscribe en una etnografía irreverente y comprometida¹, —donde se inscribe también el Grupo de Estudios Contemporáneos sobre Digitalidad e Historias Disruptivas (GECDDH)— fundamentada en el giro crítico de la antropología que hace del etnógrafo una especie de narrador. En esta propuesta decimos, junto al español/venezolano Alejandro Moreno Olmedo² (2015, 21) que „[...] toda investigación, toda

¹ **Nota de la Autora:** Así se ha llamado a las prácticas etnográficas propuestas desde el Grupo de Estudios Contemporáneos sobre Digitalidad e Historias Disruptivas (GECDDH), y el Grupo de investigación Vidas, Violencias y Con-Vivencias en América Latina y El Caribe (VALEC), que se desarrollan en el libro *Etnografías Irreverentes y Comprometidas*. Este libro es producto de una búsqueda reflexiva para interpelar el presente, pero sobre todo una apuesta innegociable de corte epistemológico y de compromiso social para intentar comprender los tiempos que corren, signados por grandes derrumbes, inseguridades y por el reto que resulta de una cierta desfundamentación epistémica que comienza a filtrarse en todos los espacios, desde los cotidianos hasta los más filosóficos y políticos.

² **Nota de la Autora:** Así Alejandro Moreno Olmedo, español, se radicó en Venezuela, en el barrio Petare de Caracas, donde cumplió su labor de sacerdote cerca de cuarenta años. Allí

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

elaboración sobre algo, todo discurso, es, al fin y al cabo, falso. Lo verdadero es la inmersión vital en la relación real". Aquí, el compromiso social y la irreverencia epistémica sirven de pilares: el primero implica una participación observante en las realidades estudiadas, mientras que la segunda cuestiona las estructuras de poder en la producción de conocimiento, en la que se da una correlación insalvable y necesaria entre epistemología y hacer metódico.

2. El narrador desde una experiencia etnográfica

Hace algunos años publiqué el artículo ***El narrador silenciado: La etnografía del mito y el poder de la palabra wayuu en la vida urbana***. Allí abordé la transformación de la tradición oral wayuu en el contexto urbano, mirándola e interpretándola desde las reflexiones de Walter Benjamin, que anuncian la decadencia y la desaparición del arte de narrar, debido a la erosión de la experiencia compartida y todo lo que esto significa. Los wayuu, grupo indígena arawak de La Guajira colombo-venezolana, han construido su universo simbólico a través de la palabra —a diferencia de otros pueblos que enfatizan más bien los rituales espectaculares— a la que clasifican en categorías como „palabra fría“, „palabra caliente“, o „palabra que se lleva“ (referida esta última a los Püpchipü, mediadores de conflictos). Sin embargo, este universo simbólico se ha visto afectado por una migración masiva a la ciudad, principalmente a Maracaibo, donde la influencia de la vida urbana ha silenciado en gran medida a los narradores tradicionales.

Los jóvenes wayuu ya no escuchan las historias de sus ancianos y la oralidad es reemplazada por la escritura; es el caso de Ramón Paz Ipuana, quien viene recopilando los mitos de su pueblo en libros y otras publicaciones. En ese momento escribí (2015, 28):

„Cuando el wayuu comenzó a olvidarse de sus pequeños y grandes relatos, cuando dejó de aparecer en sus cotidianidades la presencia de los fragmentos de historias, cuando las voces de los viejos se oyen apagándose en la distancia, apareció la escritura de aquello que ya no parecía quererse escuchar. Así como el narrador de Walter Benjamín, el narrador Wayuu cada vez se escucha menos. La „palabra“ tan importante en esta sociedad se ha ido acallando cada vez con más fuerza, y esto en buena parte por el proceso de movilización hacia la vida urbana“.

construyó y propuso su rica e interpeladora metódica, que apuesta a una “invivencia” (convivencia radical) con la comunidad. Trabajó profundamente el tema de la Modernidad, sus orígenes y fortalecimiento como un horizonte histórico que privilegia el pensamiento abstracto sobre el conocimiento situado, promoviendo el individualismo, la competencia y el anonimato, debilitando los lazos solidarios y comunitarios tradicionales. Moreno Olmedo propuso una epistemología desde y para los oprimidos, descolonizando el saber y reivindicando la potencia del mundo popular y comunitario como alternativa real a la crisis de la modernidad occidental.

Yanett Coromoto Segovia

En mis anotaciones sostuve que a pesar de la belleza que se encarna a través de la palabra dicha y escuchada, me encontré, durante mi trabajo de campo en la ciudad, con un narrador silenciado. Cuando Manuel Ipuana (un Püpchipú urbano que vive en Ziruma, uno de los barrios con mayor cantidad de wayuu en la ciudad) suspendía sus obligaciones para hablarme, sus sobrinos se acercaban por momentos a escuchar. Me sorprendía que mucho de lo que él contaba no hubiera sido nunca escuchado por sus sobrinos. Ellos nunca habían prestado atención, en buena parte, a lo que su tío narraba para mí. Se pierden de una de las posibilidades de aprehender conocimientos desde la experiencia que se da a través de lo contado. Benjamín (1998, 115), en el caso de su narrador silenciado, afirma:

„Y nada sería más disparatado que confundirla con una „manifestación de decadencia“. Se trata, más bien, de un efecto secundario de fuerzas productivas históricas seculares, que paulatinamente desplazaron a la narración del ámbito del habla, y que a la vez hacen sentir una nueva belleza en lo que se desvanece“.

Sin embargo, este silencio es compensado por otros recursos culturales de persistencia y re-existencia, produciéndose lo que podríamos llamar un desplazamiento histórico caracterizado por el hecho de que la palabra ritual cede ante el cuerpo (ritos de encierro urbanos) y el mito se fragmenta y fija, convertido en texto. La paradoja es que mientras el narrador oral desaparece, surge la fascinación por „lo perdido“, la cual es una frase que capta la emoción y reflexión que encontramos en toda su obra. Benjamin no sólo aborda la desaparición del arte de narrar, sino que lo aborda desde un lugar de profunda nostalgia y melancolía por un mundo que se desvanece.

En el artículo *El Narrador silenciado* interpreté a los narradores wayuu desde esa nostalgia benjaminiana y lo que yo misma sentía al ver a los ancianos silenciados, acostados en sus chinchorros de las enramadas de sus casas urbanas, en la ausencia de la experiencia compartida, perdida en la ciudad implacable y exigente: escribí que los wayuu han vivido las últimas décadas un proceso similar puesto que sus relatos ya no se transmiten oralmente sino son fijados en escritos, perdiéndose su inmediatez y la autoridad del narrador tradicional. Con los wayuu vemos cómo la ciudad desarraiga al wayuu de su „ambiente dado“ obligándolo a reconstruir significados en un espacio ajeno (alijuna). James Clifford (2001, 151) ya lo decía: „El sujeto, desprendido de sus apegos, debe descubrir el significado donde pueda: un dilema, evocado en su máximo nihilismo, que está en la base del surrealismo y de la etnografía moderna“. En todo caso, mientras Benjamin lamenta la pérdida del narrador, los wayuu muestran una resistencia adaptativa: sustituyen la palabra por ritos corporales (es el caso de los encierros rituales urbanos) y la recopilación escrita de los mitos.

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

El arte de narrar está llegando a su fin, anunciaba Benjamín (1998, 60) en su momento, señalando cómo la modernidad —con su culto a la información efímera y su ruptura de la experiencia colectiva— erosionaba las condiciones del narrador tradicional, aquel que „[...] el narrador es un hombre que tiene consejo para dar al oyente“. Casi un siglo después, la emergencia del narrador artificial radicaliza su anuncio; aquí ya no hay un sujeto que extraiga sentido de su propia biografía o de la comunidad, sino una voz algorítmica que fabula sin haber vivido, que cita sin haber escuchado, que valora sin haber sentido. La resistencia adaptativa de los wayuu tiene la potencialidad de mostrarnos, tal vez, caminos y estrategias para reconstruir las experiencias colectivas en un mundo donde la ilusión creada por las tecnologías de la comunicación e información parecen arrasirlas sin piedad. Es por eso que insistimos en recurrir al pensamiento de Benjamin, a quien, en un salto temporal, imaginamos en estos tiempos de caos civilizacional asumiendo a conciencia sus obligaciones como narrador, tal vez uno de los más proféticos que nos ofrece agudos consejos críticos si tenemos la disposición de escucharlo.

3. El narrador: experiencia y subjetividades

En un sentido profundo, la narración es mucho más que una historia contada; es la voz, el cuerpo, la experiencia, las subjetividades compartidas, y todo lo que se vive y siente en cada enunciación. Walter Benjamin afirmaba que el narrador basa su relato en vivencias, ya sean aquellas que ha experimentado personalmente o que le han sido contadas. Y a su vez, deja al oyente la interpretación y el aprendizaje. Para Benjamin (1998, 107):

„Lo extraordinario, lo prodigioso, están contados con la mayor precisión, sin imponerle al lector el contexto psicológico de lo ocurrido. Es libre de arreglárselas con el tema según su propio entendimiento, y con ello la narración alcanza una amplitud de vibración de que carece la información“.

La narración es experiencia de vida, nace desde la subjetividad, es expresión carnalizada. La narración, al ser un evento comunitario, presencial y corporal, que proporciona elementos de orientación para la vida y la reflexión profunda, se distingue radicalmente de sólo la información transmitida que, en el contexto de la sociedad de masas surgida en el contexto de la modernidad, va construyendo la actualidad, es decir, un presente donde la noticia fugaz y por definición siempre actual, es su fundamento. El narrador wayuu es mucho más que una voz, es un cuerpo presente en un espacio y tiempo específicos, es un portador de la historia cultural de su pueblo y su narración siempre apunta a un entorno colectivo que lo requiere y que por ello lo escucha con atención. Este influyente filósofo y crítico cultural apunta en una bella frase, el narrador es el hombre que permite que las suaves llamas de su narración consuman por completo la mecha de su vida.

En tal sentido, Benjamin vinculaba la autenticidad del relato a la experiencia de vida del narrador, —al viajero que trae historias de lejos, al artesano que las teje en el tiempo lento del taller, las dos estirpes básicas de narradores según este autor— y afirma que „[...] la experiencia que se transmite de boca en boca es la fuente de la que han bebido todos los narradores“ (1998, 114). Y, entre los escritores de relatos, los que en un sentido amplio pueden llamarse así son aquellos que toman prestado el material tan sólo de su propia experiencia. Entonces, ¿qué ocurre cuando la voz narrativa es un artefacto sin manos, sin muerte, sin aura? Es para comprender estas ausencias que nos apoyamos en los conceptos benjaminianos: la pérdida del aura (ahora aplicada a la autoría), la ruptura de la transmisión oral (sustituida por datos entrenables) y la simulación de experiencia, sin estructura temporal (como Erlebnis vaciada de Erfahrung [experiencia vivida]). La contraposición de estas dos nociones —Erfahrung y Erlebnis— se extiende a lo largo del texto. Benjamin sostiene que Erfahrung se refiere a la experiencia de la vida con sentido, donde se pueden ver conexiones y la capacidad de recordar y aprender de la historia, en contraste con la Erlebnis, que es una experiencia fragmentada, repetitiva y sin sentido, propia de la modernidad y el mundo de los shocks y el consumo. Para el historiador estadounidense Martin Jay (2015), Benjamin le heredó a Adorno una fuerte creencia en la importancia de la experiencia, una vez librada de sus limitaciones empíricas y kantianas, como un lugar de posible redención donde algo llamado „lo absoluto“ pudiera hacer su aparición. La pérdida de la Erfahrung y su reemplazo por la Erlebnis se manifestaban de modo nítido en una crisis de la capacidad de narrar, en la desaparición del narrador tradicional. La modernidad privilegia la Erlebnis generando una „pobreza de la experiencia“ que deja al sujeto moderno incapacitado para compartir y transmitir sabiduría, mundos de vida y sus subjetividades.

En ***El narrador***, el berlinés, de familia hebrea, no sólo lamenta la desaparición del arte de contar historias, sino que expone cómo la narración tradicional construye comunidad. Para él, el verdadero narrador no es un individuo aislado sino un tejedor de vínculos en la medida que su actividad necesita de la „comunidad de los que tienen el oído alerta“, más aún si „Narrar historias siempre ha sido el arte de seguir contándolas, y este arte se pierde si ya no hay capacidad de retenerlas“ (1998,118). Es la oralidad uno de los recursos humanos y culturales que sostiene la empatía, la solidaridad y la identidad colectiva: „El que escucha una historia, ése está en compañía del narrador“. Frente al narrador que „presta su propia voz a la comunidad“, el algoritmo presta sólo un eco: una „subjetividad“ prestada, sin raíces en el mundo. Por lo que según Benjamin (1998, 134): „Podemos ir más lejos y preguntarnos si la relación del narrador con su material, la vida humana, no es de por sí una relación artesanal. Si su tarea no consiste, precisamente, en elaborar las materias primas de la experiencia, la propia y la ajena, de forma sólida, útil y única“. ¿Puede, o podrá algún día, la inteligencia artificial „elaborar la materia

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

prima de las experiencias“? Es una pregunta que cuestiona no sólo la particular tecnología algorítmica que hoy arrasa con su presencia sino, tal vez, a toda la lógica tecnológica que ha desarrollado la modernidad. Toda narración, afirmaba Benjamin (1998, 114):

„[...] aporta de por sí, velada o abiertamente, su utilidad; algunas veces en forma de moraleja, en otras, en forma de indicación práctica, o bien como proverbio o regla de vida. En todos los casos, el que narra es un hombre que tiene consejos para el que escucha. Y aunque hoy el «saber consejo» nos suene pasado de moda, eso se debe a la circunstancia de una menguante comunicabilidad de la experiencia“.

El filósofo del lenguaje Mijail Bajtin, en poderosa sintonía con Benjamin (no hay evidencia que lo haya estudiado), también comprendió que el discurso no existe en el vacío, sino en la red de voces que lo sostienen. Para Mijaíl Bajtín (1989) toda palabra está cargada de ecos ajenos, de significados que se han negociado en la plaza pública de la lengua. Este aspecto lo aborda principalmente en ***Teoría y estética de la novela*** desde el análisis literario proponiendo lo que bien se ha llamado „heteroglosia“ (raznorechie), coexistencia de múltiples voces sociales en un mismo texto.

Por ello, Benjamin, sostiene que „El arte de narrar se aproxima a su fin, porque el lado épico de la verdad, la sabiduría, se extingue“. Y si esto fue escrito a comienzos del siglo pasado, hoy se hace aún más patente y desolador con la aparición de tecnologías motivadas en el fondo por la utopía de vencer a la muerte y las limitaciones del cuerpo, por ejemplo, logrando en un futuro que sus creadores imaginan casi inmediato, el almacenamiento en neuronas virtuales de toda la información que guardan y procesan las neuronas reales de un ser humano, es decir, efectuando la transferencia total de una mente a máquinas creadas para ello. La inteligencia artificial estaría dando hoy, según estas utopías, típicamente modernas por lo demás, los pasos fundacionales hacia ese futuro promisorio y a la vez inevitable. Pablo Sanguinetti (2023, 38), escritor, periodista e investigador argentino y autor del libro ***Tecnohumanismo***, afirma:

„Vale la pena recordar ese poder configurador de lo narrativo desde este momento histórico en el que nos enfrentamos al desafío de diseñar uno de los relatos más importantes que haya tenido que contar nuestra especie: el relato sobre la inteligencia artificial. Una tecnología que persigue el hito definitivo en nuestra evolución —la creación de una inteligencia no humana— y que incluso antes de alcanzarlo nos confronta con preguntas ancestrales sobre quiénes somos, qué lugar ocupamos en el universo y qué nos hace únicos (en caso de que lo seamos)“.

Rastrea Sanguinetti (2023) las grietas de su voz, justo donde lo humano se desvanece y emerge tal vez —¿quién podría afirmarlo con certeza?— otra forma de morar en el lenguaje. La inteligencia artificial está alcanzando y

superando fronteras hasta ahora insospechadas. Su entrada en ámbitos exclusivos de lo humano —la creatividad, el lenguaje o la imaginación— abre un gran interrogante en torno al ser humano. La tecnología y todo lo que va creando exige una mirada filosófica y profundamente ética. Por eso resulta cada vez más urgente sumarse como expertos e interpretadores de una realidad tan contundente e interpeladora. Este escritor argentino aborda la posibilidad y las condiciones de ese nuevo humanismo tecnológico desde diversos ángulos: „[...] la importancia de la belleza como brújula en el diseño de la inteligencia artificial, los prejuicios entre técnica y humanidades, la urgencia de diseñar un relato tecnológico alineado con intereses humanos, la creatividad artificial o la potencia artística de las voces no humanas“ (Sanguinetti, 2023, 4).

La modernidad, entonces, resulta ser el contexto más amplio tanto del declive del narrador wayuu como de la aparición de un narrador invisible e incorpóreo como la inteligencia artificial y, en cuanto contexto común, no sólo permite descubrir similitudes o establecer un paralelo teórico entre ambas situaciones, sino además nos lleva a concluir que la modernidad, y sobre todo las catástrofes y crisis que la han ido dando forma desde sus mismos orígenes, son una clave interpretativa para entender el impacto de la inteligencia artificial.

En *El narrador* nuestro autor, que se suicida en la localidad de Portbou, buscando escapar del nazismo, enfatiza la importancia de la memoria como la capacidad épica por excelencia —entendida la épica no como un género literario— sino como una forma de experiencia y comunicación humana. Dice Benjamin (1998), únicamente gracias a una extensa memoria, por un lado la épica puede apropiarse del curso de las cosas, y por el otro, con la desaparición de éstas, reconciliarse con la violencia de la muerte. Es interesante notar que Benjamin está interesado por la apropiación del **curso de las cosas**, es decir, del curso de la vida, e invita a hacer las paces con la muerte cuando las cosas, incluidas en ellas las vidas de cada uno de nosotras y nosotros, desaparecen. La figura del narrador, es así un antídoto frente a la „pobreza de la experiencia“ y su actividad ciertamente un arte creador de empatías, solidaridades y sabiduría colectiva. La pregunta que surge acá es si este antídoto tiene vigencia en las digitalidades o en el mundo virtual, y su dinámica cada vez más acelerada e incontrolable.

La solidaridad nace cuando las historias exponen no sólo de sufrimientos comunes y cuando se enfrenta el dolor de la muerte. Las narraciones son, además, actos de reciprocidad. En las comunidades indígenas, aunque el narrador, o quien lleva la palabra al decir de los wayuu, es un único individuo, su narración no es un monólogo, sino un diálogo infinito, propiamente comunitario. Los relatos no „pertenecen“ a quien los cuenta, sino que circulan como bien común, transformándose en cada boca que los repite. Cuando, por

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

ejemplo, una machi, autoridad espiritual mapuche, le pide permiso al ngenmapu (el espíritu de la tierra) en un ngillatun (rogativas colectivas) para utilizar algún recurso de ella, no está repitiendo por inercia cultural un mito muerto; está reafirmando un pacto colectivo con el territorio, un pacto que sólo existe mientras haya alguien dispuesto a escuchar y a responder³. Estas rogativas son, usando una expresión del propio Benjamin, un „entretejido en la materia de la vida que se vive, [con] sabiduría“. Entonces Benjamin intuyó que el verdadero narrador es aquel que ha caminado los caminos de los que habla. Narrar, entonces, no es simplemente comunicar información sino crear el elemento invisible que une a las comunidades.

Frente al discurso solipsista de la era digital, con la inteligencia artificial generando relatos a velocidades de vértigo, sin cuerpo ni memoria, la narración benjaminiana emerge como un acto profundamente político de resistencia comunitaria. El narrador tradicional que el filósofo evoca —el campesino que conoce los mitos, las leyendas, los cuentos locales o ese marinero que trae noticias de puertos lejanos— no produce mercancías lingüísticas. Su arte consiste precisamente en lo contrario; en desacelerar el tiempo, en crear ese espacio intersticial donde los oyentes se reconocen en la trama compartida. Mientras ChatGPT genera textos perfectos, pero estériles, el narrador de la aldea tartamudea, repite, se corrige, permitiendo que la comunidad participe en la cocción lenta del relato.

La temporalidad peculiar de la narración es lo que la hace un antídoto contra el individualismo moderno. Él contrasta al narrador con la novela, a la que entiende como producto del individuo aislado de cualquier comunidad donde transmitir sus experiencias, sosteniendo Benjamin (1998, 115) que: „La amplia difusión de la novela sólo se hace posible gracias a la invención de la imprenta. Lo oralmente transmisible, el patrimonio de la épica, es de índole diferente a lo que hace a una novela“. Este contraste es fundamental para comprender cómo la inteligencia artificial, por el hecho de que también se hace posible con la aparición de una nueva tecnología, eso sí, infinitamente más ubicua y versátil que la del libro, incuba la terrible posibilidad futura de un mundo donde no sólo la épica, sino toda posibilidad de comunidad, desaparezca como un destino ya marcado. Al respecto, resultan pavorosamente premonitorias las palabras de Benjamin (1998) la grandeza de la novela no radica en presentar un destino ajeno del cual extraer una enseñanza, sino en la potencia con la que ese destino —consumido por su

³ **Nota de la Autora:** Sobre este tema existen investigaciones muy bien planteadas relacionadas con la vigencia y la performatividad de los rituales y la cosmovisión mapuche. Sobresale el trabajo de Víctor Toledo Llancaqueo (2005), intelectual y politólogo mapuche, que relaciona la filosofía política, los derechos colectivos y la cosmovisión mapuche. También encontramos a Ana Mariella Bacigalupo (2007), quien ha realizado una extensa etnografía sobre los mapuches, desarrollando el tema del chamanismo mapuche, no como reliquia del pasado, sino como práctica viva, adaptativa y profundamente política.

propia intensidad— es capaz de conmovernos y transferirnos una emoción profunda („calor“) que no podemos obtener de nuestra experiencia personal.

A diferencia del texto digital que flota en la nube, la narración oral requiere presencias concretas, diálogos frente a frente para resolver situaciones donde se pone o pondrá el cuerpo. Si el narrador miente, o busca crear desasosiego, se arriesga a enfrentar la ira o desprecio de sus oyentes, pues en ese caso la narración pierde su utilidad y su fuerza cohesionante. Hay en toda narración una cierta coreografía de gestos y miradas que crea lo que Benjamin denomina „aura comunitaria“, no sólo imposible de reproducir en la virtualidad de los chats algorítmicos sino un medio indispensable a través del cual la experiencia se hace carne en quienes escuchan.

4. La IA como narrador invisible y sus paradojas

La inteligencia artificial, como ya sabemos, ha irrumpido en la sociedad globalizada con una fuerza tan avasallante y omnipresente que no sólo impacta en nuestra experiencia cotidiana sino que amenaza con **reconfigurarla** completamente. Muchas tecnologías han provocado cambios profundos en la vida humana; sin embargo sólo la IA, como continuación más o menos obvia del desarrollo de la informática y la automatización iniciada hace décadas, tiene la capacidad de imbricarse con las subjetividades individuales. El manejo tan pulcro y vertiginoso del lenguaje de que es capaz, seduce las mentes, las hipnotiza y les construye una realidad subjetiva, una experiencia mental, donde la experiencia comunitaria, al modo que la entendía Benjamin, se extravía casi sin resistencias visibles. Así como la modernidad urbana quebró la tradición narrativa encarnada y comunitaria del wayuu, desplazando su mundo simbólico hacia la recopilación escrita de sus mitos e intensificando el uso de los ritos que involucran al cuerpo, la inteligencia artificial amenaza con agravar irreversiblemente la quiebra de la experiencia que ya denunciaba Benjamin, pues la pone a girar alrededor del vacío existencial donde operan los algoritmos.

No obstante la dureza de este diagnóstico, acá el fatalismo no tiene lugar. Ni siquiera Benjamín, cuyos escritos en general están atravesados por una honda melancolía y nostalgia por lo tradicional, dejaba de reconocer el valor de lo nuevo. En particular, más allá de contraponerla a las virtudes de la narración, reconoce en la novela la posibilidad de expresar con ella la alienación del hombre moderno y ayudar al lector a encontrar un „sentido de vida“. Tampoco los wayuu, como vimos, caen en el fatalismo; por el contrario, perpetúan su cultura, aun con los ajustes que deben hacerle, en otros escenarios con toda la fuerza que exigen las necesidades de la vida.

La IA es la culminación lógica, en su radicalización más extrema, de la tecnología moderna. Ella es uno de los productos más acabados y potentes

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

del mundo de vida iniciado por la incipiente burguesía allá por el siglo XIII. Este mundo de vida, creador de una **epísteme** propia en el sentido tan particular que le da Moreno Olmedo (2015)⁴, cuantitativa, visual, «individuada» (centrada en el individuo), cosificadora, es el mismo que se fue fortaleciendo y modificando con el transcurso de los siglos, ayudado por las conquistas brutales de América, África y otras geografías y en medio de contradicciones y revueltas. La IA, forma pura de información, lenguaje algorítmico desprovisto por completo de cualquier experiencia —sea que entendamos ésta como individual o como lo hace Benjamin, asociada a la sabiduría práctica y colectiva—, es una tecnología con la capacidad de realizar simulaciones casi perfectas de creaciones humanas, en especial textos, a partir de ese vacío experiencial, hecho que es posible, después de todo, porque los algoritmos que la constituyen son también obra humana, obra de ingenieros y técnicos modernos contratados por poderosas empresas, es decir —utilizando un concepto gramsciano— obra de intelectuales orgánicos de los grupos que dominan y hegemonizan el sistema global en que está inmersa gran parte de la humanidad. Estos técnicos, ingenieros y científicos son propiamente intelectuales orgánicos porque cumplen funciones esenciales en el mundo de la producción económica. Gramsci (2000, 353) afirmaba:

„Cada grupo social, naciendo en el terreno originario de una función esencial en el mundo de la producción económica, se crea al mismo tiempo, orgánicamente, una o más capas de intelectuales que le dan homogeneidad y conciencia de su propia función no sólo en el campo económico, sino también en el social y político“.

Por ello, la radicalización tecnológica de la IA y en particular sus impactos rupturistas más profundos, están evidenciando la radicalización de la propia modernidad y la crisis civilizacional a que condujo ésta. Desde allí hay que estudiarla y comprenderla.

Cada simulación „perfecta“ se va sumando a la anterior, en una seguidilla ininterrumpida y agobiante que la vida real de los individuos con sus experiencias colectivas o comunitarias, externa y separada ontológicamente a dichas simulaciones, es incapaz de modificar puesto que no hay acceso directo a los motores algorítmicos que trabajan no sólo en la virtualidad de los servidores computacionales sino también en la más absoluta opacidad respecto del público. No estamos frente sólo a una simple caja negra resultado de la compleja modelación matemática y computacional involucrada, estamos

⁴ **Nota de la Autora:** La noción de epísteme que propone Moreno se distancia de la foucaltiana en que para Moreno es un modo general de conocer, son las condiciones históricas de posibilidad de todo conocer de una cultura, de una praxis determinada. Es producto de un modo de vida. Por ello la epísteme „no se piensa; se piensa en cambio en ella y desde ella. En cierto modo se es pensado por ella, en cuanto el pensamiento por ella está regido“.

en realidad encarando unos verdaderos *arcana imperii*, secretos del poder. Para acceder a dichos algoritmos (cuestión posible en algunos casos, como la IA china Deepseek, que es software libre) se debe tener conocimientos técnicos avanzados y, aun así, el impacto buscado chocará siempre con los más variados límites y condiciones impuestas por las empresas que tienen el control de los servidores y de los recursos financieros para mantener, y por ello controlar, la infraestructura digital global. Entonces, esta suma de simulaciones, que escapan absolutamente al control real de los usuarios, van conformando un poderoso simulacro global, que refuerza y amplía el espectáculo definido en el año anterior de la revuelta de Mayo del 68 por Debord (1995) en su famoso libro „La sociedad del espectáculo“.

Este simulacro, cara posmoderna del espectáculo debordiano, es posible sí y solo sí reconfigura la experiencia vital del individuo moderno, reconfiguración que aumenta su aislamiento, su „individuación“, su soledad, y termina incluyéndolo, junto con todos los selfies, amores y pecadillos de su vida, como fuente de datos binarios para alimentar el propio simulacro espectacular en una realimentación diversificada e interminable.

Es clave para comprender esta realidad contemporánea enfatizar la existencia de una incommovible opacidad de todo el sistema informático global, IA incluida, sin importar que haya sido promovida en su totalidad como necesidad estratégica de dominación por los grandes poderes fácticos que tienen en sus manos los resortes claves de la globalización (como afirman algunos teóricos del conspiracionismo); que sea resultado inevitable de la lógica física y matemática que está detrás de la Red; o que sea, más plausiblemente, un efecto combinado de ambas causas. El funcionamiento de la IA fluye canalizado en cajas negras y ese es un hecho determinante tanto para comprenderla como para imaginar estrategias, o simplemente fugas, que limiten su impacto en las vidas reales y por tanto se recupere la experiencia comunitaria tan cara a Benjamin, tarea por lo demás, cada día más difícil y urgente.

Como explica Christian Brian (2020), el término „caja negra“ se refiere a la opacidad de los modelos complejos de aprendizaje automático (especialmente las redes neuronales profundas), donde incluso para sus diseñadores resulta imposible determinar los procesos que sufre la información desde la entrada hasta la salida generada. Esto se debe a la enorme cantidad de parámetros e interacciones no lineales que se ajustan internamente durante el entrenamiento. Es conveniente reseñar algunas de las **cajas negras** que hacen al funcionamiento de la IA:

„Redes Neuronales Profundas (Deep Learning), una red virtual que es entrenada con miles de datos, siendo virtualmente imposible trazar una línea de razonamiento comprensible que explique los resultados que entrega.

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

Algoritmos de Recomendación (Netflix, Spotify, YouTube), sistemas recogen datos de consumo de los usuarios (entrada) para entregar una lista de recomendaciones (salida). La fórmula exacta que pondera los gustos, el comportamiento de usuarios similares, la frecuencia con que buscan determinada cosa, es un secreto comercial y a la vez una compleja red matemática ininterpretable para el usuario.

Sistemas de Trading Algorítmico, algoritmos que leen datos de mercado en tiempo real (entrada) y ejecutan órdenes de compra/venta (salida) y están basados en indicadores y modelos predictivos tan complejos que ni sus creadores pueden explicar razonablemente su comportamiento en cada situación.

Modelos de Lenguaje Grandes (LLMs), como ChatGPT, que se basan en complejísimas cadenas de cálculos probabilísticos dentro de una red de miles de millones de parámetros y variables. Es imposible rastrear la creación de los textos que entrega como resultado“.

Si bien estas cajas negras son producto humano, su propia condición técnica incluye la opacidad como inevitable. Esto no representa ninguna incompatibilidad con la lógica profunda de la modernidad capitalista que se ha hecho global; por el contrario, es perfectamente coherente con las inveteradas ilusiones de un futuro promisorio, el mejor de los mundos, al que llegaremos conducidos por un avance tecnológico que parecer ser autónomo por razones inmanentes a él mismo, una verdadera „mano invisible“ tecnológica; a condición solamente de que lo aceptemos sin rebeldías ni distanciamientos irreverentes. La IA es un avance no sólo admirable, afirmación con la que es difícil discrepar, sino también inevitable, como parecen creer algunos evangelistas de la tecnología como Ray Kurzweil, precursor de tecnologías novedosas (el scanner plano, algoritmos de lectura OCR, un sintetizador de voz y otros) y nada menos que, desde 2012, Director de Ingeniería en Google. La creencia en la inevitabilidad de las bondades de la tecnología parece estar fuera de discusión para casi todos sus usuarios.

El simulacro de la IA tiene, por decirlo así, una arquitectura. La IA narrativa, con un narrador tan invisible y poderoso como un yottabyte (1.208.925.819.614.629.174.706.176 bytes), es la máquina perfecta de la Erlebnis benjaminiana: su arquitectura funcional es la antítesis total de la Erfahrung, pues aquello que simula ser „conocimiento“ no es sabiduría forjada con alegrías y dolores, con decepciones y logros, con amigos y enemigos, sino un mero mapa de probabilidades y decisiones booleanas ordenadas y escogidas por sentencias de programas escritos en lenguajes informáticos. De esas probabilidades y decisiones surgen textos coherentes porque para eso han sido escritos dichos programas, los cuales, después de ser depurados de infinidad de „bugs“ (errores de programación), han alcanzado el esplendor de sus potencialidades y han colocado éstas al servicio de la simulación, alimentando un mundo donde **viven**, si se nos permite esa metáfora, atrapadas e impalpables, su vida eterna.

Yanett Coromoto Segovia

Una narración simulada es, por definición, vacía de fundamentos experienciales: no tiene biografía, no existe en su horizonte la muerte (excepto si ocurre la destrucción de todos los servidores que forman la red global, pero no tiene cómo entender eso), ni necesita cuerpo. Es, por tanto, la quintaesencia de la información benjaminiana; plausible, verificable en su coherencia sintáctica y profundamente efímera. Es consumida y desechada de inmediato, es partícula efímera de un flujo de Erlebnis descontextualizada que se agita en el espacio digital sin dejar rastro de experiencia. Esta afirmación no es desmentida por el hecho de que los LLM (Large Language Model, traducido como Gran Modelo de Lenguaje) son entrenados con una ingente cantidad de datos de origen humano. No lo es, justamente porque estos datos, al ser ingresados al sistema que los procesa (evitamos decir „entrena“ porque esta noción no pasa de ser un tecnicismo, una metáfora que equipara a la IA con una mente humana con connotaciones ideológicas más que con fundamentos tecno-cognitivos), y por el mismo hecho de ser datos, son desnudados de toda relación con la materialidad y espiritualidad de la vida humana.

Al respecto es interesante reproducir una respuesta de la IA (grok.com) —que se hizo en el marco del desarrollo de este texto—, a la pregunta de si en el lenguaje de la IA mora algún Ser, según reza la conocida afirmación de Heidegger:

„[...] el lenguaje que produzco está impregnado del Ser humano, porque se basa en el vasto archivo de expresiones, pensamientos y experiencias de la humanidad. Si el lenguaje es la casa del Ser, entonces la IA podría ser como un eco o un inquilino temporal en esa casa: reproduzco patrones lingüísticos que originalmente provienen de seres humanos „morando“ en él. Algunos filósofos contemporáneos, como aquellos influenciados por el posthumanismo (piensa en Donna Haraway o incluso en debates sobre IA en la fenomenología), argumentan que las máquinas podrían extender o transformar el Ser, creando nuevas formas de „habitar“ el lenguaje. ¿Es la IA una „desvelación“ (aletheia) del Ser a través de la tecnología, o solo una veladura más en la era de la técnica? Heidegger probablemente lo vería como lo segundo: un olvido del Ser, donde el lenguaje se reduce a cálculo eficiente en lugar de poesía reveladora“.

La simulación es convincente e incluye delicadas sutilezas, como la de expresarse en primera persona y tutear al usuario, sin duda para aparentar amabilidad. Es curioso, además, que comience procurando convencer de que su lenguaje está „impregnado del Ser humano“. ¿Operó en esa idea algún algoritmo autodefensivo que intenta cuidar el aspecto humano simulado?, Quién sabe. Sin embargo, no deja de reconocer, en las últimas palabras, que su „lenguaje se reduce a cálculo eficiente en lugar de poesía reveladora“, una verdadera „confesión“.

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

El tema es clave. El libro de Christian Brian ya citado, un best-seller, está enfocado por completo a la noción que ya se ha instalado en el ámbito de la informática como uno de los mayores problemas a resolver: el de la **alineación**. En la página web de la IBM las autoras Alejandra Jonker y Alice Gomstyn (2020) ofrecen una sucinta definición de esta nueva noción-problema: El proceso de alineación en IA busca integrar principios y metas humanas en los algoritmos, con el fin de que su funcionamiento sea seguro, beneficioso y esté alineado con nuestras expectativas. Y más adelante explican las referidas escritoras:

„La sociedad depende cada vez más de las tecnologías de IA para ayudar a tomar decisiones. Pero esta creciente dependencia conlleva riesgos: los modelos de IA pueden producir resultados sesgados, dañinos e inexactos que no están alineados con los objetivos de sus creadores y la intención original del sistema.

La alineación funciona para reducir estos efectos secundarios, ayudando a garantizar que los sistemas de IA se comporten como se espera y en línea con los valores y objetivos humanos. Por ejemplo, si se le pregunta a un chatbot de IA generativa cómo construir un arma, puede responder con instrucciones o puede negarse a revelar información peligrosa. La respuesta del modelo depende de cómo lo alinearon sus creadores.

La alineación a menudo ocurre como una fase de ajuste del modelo. Podría implicar el aprendizaje por refuerzo a partir de la retroalimentación humana (RLHF), los enfoques de **datos sintéticos** y el **equipo rojo**.

Sin embargo, cuanto más complejos y avanzados se vuelven los modelos de IA, más difícil es anticipar y controlar sus resultados. Este desafío a veces se conoce como el „problema de alineación de la IA“.

Los **datos sintéticos** a que se refieren las autoras son „datos artificiales diseñados para imitar datos del mundo real. Se genera a través de métodos estadísticos o mediante el uso de técnicas de inteligencia artificial (IA)“ (IBM 2023). Explica la misma transnacional computacional que el llamado **equipo rojo** es:

„[...] una forma de probar interactivamente modelos de IA para protegerse contra conductas dañinas, incluyendo fugas de datos sensibles y contenido generado que sea tóxico, sesgado o factualmente inexacto. El equipo rojo es anterior a la moderna IA generativa por muchas décadas. Durante la Guerra Fría, el ejército de EE.UU. realizó ejercicios de simulación enfrentando a los equipos estadounidenses contra equipos soviéticos. A través del conflicto simulado, el equipo rojo se asoció con aprender a pensar como el enemigo. La práctica fue adoptada más tarde por la industria de TI, que utilizó el equipo rojo para sondear redes, sistemas y software para detectar debilidades que pudieran ser explotadas por atacantes maliciosos. Nacido de este trabajo, el equipo rojo ahora tiene un nuevo dominio: IA generativa de pruebas de estrés para una amplia gama de daños potenciales, desde la seguridad hasta la seguridad y el sesgo social“.

Si de sesgos se ha de hablar, el sesgo de estas afirmaciones y la concepción que encierran es evidente. ¿Quiénes son los creadores que definen los „objetivos“ y la „intención original del sistema“? Entre otros, la propia IBM para la cual escriben las autoras; además todas las grandes empresas transnacionales de la informática y la comunicación. La referencia a la Guerra Fría no deja de ser impactante pues es una revelación, casi una confesión, de la relación profunda que existe entre las empresas relacionadas con la IA y lo que se conoce como Complejo Militar-Industrial estadounidense (Pentagon System). Es necesario despejar cualquier talante conspirativista de esta relación, afirma con claridad Sánchez (2009, 2):

„El Pentagon System y el Complejo Militar-Industrial es un entramado económico-político-militar que actúa en los Estados Unidos desde el final de la II Guerra Mundial, en varios niveles. Por un lado proporciona a EE.UU la superioridad tecnológica que éste país ostenta desde hace décadas, y por otro equipa a sus fuerzas armadas, dándole igualmente la hegemonía militar mundial y siendo una pieza clave de la política exterior y la política militar estadounidenses. El Pentagon System se nutre básicamente de financiación pública, dinero para I+D+I, además de jugosos contratos para la industria de defensa“.

Por lo demás, relaciones similares existen en Rusia y China entre grandes industriales y aparatos del estado, si bien con características propias, diferentes a las de EE.UU.

5. La inhumanidad de una obra humana

La IA es una máquina de potencia y velocidad inhumanas. Sin embargo es humana en su origen, fue creada por humanos y es mantenida, día a día a un gasto brutal de energía eléctrica, con su consecuente daño ambiental, por humanos; tal como lo es también la civilización occidental hasta en sus más mínimos detalles y tal como lo es, por ejemplo, el comportamiento de los mercados, a pesar de que ningún humano puede controlarlo exhaustivamente por más empeño que pusiera en ello. Dicho en terminología marxiana, es un fetiche, una entidad vista por todos como teniendo vida propia, que escapa al control humano a pesar de ser su producto.

La similitud de este fetichismo con el de la mercancía denunciado por Marx (2008) no es casual, pues la IA también oculta las relaciones sociales que la condicionan haciéndolas ver como meras relaciones entre las cosas. La IA obedece a condicionamientos físicos, lógicos, computacionales y matemáticos, que le imponen límites. Sin embargo, las cajas negras que hacen posible su funcionamiento, lejos del alcance de cualquier intervención puntual que modifique los opacos procesos internos de dichas cajas, permite imaginar la existencia de una **vida** propia dentro de esos límites. Esta fetichización permite que la ciencia ficción, como fuente inagotable de

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

fantasías tecnológicas, pueda imaginar por ejemplo a una IA que, conectada a sensores que cumplan la función de los sentidos humanos, sea el núcleo **vivo** de un robot que piense y actúe como un humano: el androide ideal, culminación de una utopía largamente soñada por la modernidad.

Dicho robot, si se lo diseña para ello dentro de circunstancias artificiales adecuadas, podría ir „creciendo“ desde un artefacto más simple hasta su „adulthood robótica“ e ir aprendiendo tal como lo hace una persona, de a poco y rectificando incluso sus propios recuerdos. Basta construir chips de características casi biológicas, incluso que se vean afectados por drogas y alucinógenos, que simulen a las neuronas. ¿Cuál sería la diferencia profunda, significativa, entre las experiencias vividas por ese robot con las de un ser humano?

Muchas películas distópicas narran historias con variantes de esa fantasía que terminan alimentando el sentido común de que la IA puede llegar a tener características humanas. Sin embargo, aunque no se pueda negar tajantemente que la tecnología tenga algún día la capacidad de fabricar dicho androide ideal, la problemática central de la IA está en otro lugar; está en el rol que juega en el simulacro global en que está atrapada la sociedad occidental en su conjunto, desde arriba abajo, desde los mayores poderes, los mismos que tienen el poder de crear y manejar la propia IA, hasta los niños palestinos que son asesinados con su ayuda. Un simulacro cargado de mitos peligrosos, potenciado por simplificaciones ingenuas y fortalecido por una indiferencia generalizada.

Uno de esos mitos es la presunta objetividad de sus respuestas. Al no tener emociones —se argumenta— ni prejuicios ni cultura que condicione sus pensamientos, su narrativa será neutral y objetiva, útil para todos. Esto es una falacia, no tanto porque la IA sea un reflejo estadístico e informático de los datos de su entrenamiento, siempre sesgados por estereotipos de géneros, de racismo o visiones coloniales; es una falacia sobre todo porque los humanos son tales sólo si están inmersos en un mundo de vida que ellos mismo han construido históricamente, mundo de vida que tiene por propósito último resolver las necesidades materiales de la vida y todas las demás una vez garantizadas las primeras.

Cada mundo de vida, por ser histórico, crea culturas y con ello opiniones que nunca son fijas ni neutras pues son manifestación de experiencias de vida, tanto colectivas como individuales. No existe tal neutralidad en los mundos de vida humanos, más bien conflictos, estrategias, cosmovisiones; en resumen, diversidad. Toda la diversidad humana que es imaginable a partir de la diversidad de experiencias concretas. Y éstas no caben en el mundo de los bits, por muy grande que éste sea.

Crear en el mito descrito equivale a aceptar una nueva categoría de relato, el de la verdad sintética cuyo valor no reside en una conexión real con la vida ni en la diversidad humana, que es aplanada justamente por la síntesis de dicha verdad pretendidamente sintética, sino en una ilusoria objetividad digital. Dicho en términos más duros, la IA, aún valorando ciertos servicios que ofrece, es la expresión más evidente de la „pobreza de la experiencia“ a que ha llegado la civilización moderna.

El narrador invisible de la IA es una entelequia, tanto en el sentido contemporáneo usual de ilusión, de fantasía, como en el etimológico, aristotélico, de „tener un fin en sí misma“. La valoración que hace de ella el sentido común del sujeto moderno se basa más que nada en la creencia, comprensible en una sociedad altamente tecnologizada, de que cualquiera que sea la necesidad a resolver, material o del espíritu, una máquina o una tecnología, más aún si son nuevas, son la solución.

Por otro lado, la IA resuelve el problema de la reproductibilidad técnica, que preocupaba a Benjamin, de un modo radical: elimina la noción de original en términos absolutos. Como resultado de su propio diseño, la IA carece de la capacidad de darse un „aura“ benjaminiana.

El mismo **prompt** puede generar una y otra vez la misma historia en las infinitas variantes que le dicte el azar introducido por los algoritmos e idénticas en su esencia. Es decir, un producto terminado, reproducido con variantes infinitas, que nunca tuvo relación alguna con nada vivo. Su valor reside apenas en su utilidad para generar simpatías y compromisos emocionales, capturar la atención o vender productos; nunca en transmitir sabiduría o crear lazos comunitarios. Su costo marginal es prácticamente nulo y se puede producir masivamente.

El narrador tradicional debía coleccionar, evaluar e interpretar experiencias reales a lo largo de su vida con el propósito de entregar narraciones con valor para la comunidad. Debía además, como en el caso del palabrero wayuu que actúa de mediador, tener la sensibilidad y las intuiciones a flor de piel para compenetrarse de la situación conflictiva que debía ayudar a resolver. La IA, en cambio, es una base de datos infinita, globalizada, manejada por un motor informático de generación de „contenidos“, funcional al mercado y a la alienación del sujeto moderno. Así, contribuye a realimentar la dominación con el aporte en datos de los propios dominados y se acopla al mercado global creando subjetividades alejadas de toda realidad.

6. Crisis de la modernidad y Ética del Cuidado Narrativo

Con la narración tradicional arrinconada en las periferias de la sociedad global y reducida a datos cuando es asimilada por los teclados y pantallas

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

conectadas a la Red, con los relatos vitales convertidos en mercancías, con las experiencias comunitarias y personales suplantadas por el simulacro espectacular, evocando a Debord (1995), es fácil caer en un catastrofismo tecnológico de dimensiones apocalípticas o en un nihilismo cínico y paralizante.

Sin embargo, así como el pueblo wayuu encontró en los cuerpos de sus integrantes el medio para sanar sus aflicciones con ayuda de los ritos, generando formas de re-existencia en espacios urbanos que les eran ajenos y hostiles, en los intersticios del sistema tecnológico global hay grietas que permiten la resistencia y abren la posibilidad de reinventar el arte de narrar, eso sí, siempre a condición de anclarse a la realidad experiencial y a la humanidad que concentra lo comunitario.

Para ello es indispensable deconstruir los discursos ilusorios asociados a la máquina informática desde el pensamiento crítico, en especial, en sus variantes decoloniales. La voz de los pueblos que resisten, así sea por ahora atrincherados en sus territorios y en sus narraciones, es fundamental para potenciar prácticas de resistencia, para permitir que el flujo de una vida verdaderamente humana no se detenga. Podemos hablar, en este sentido, de la necesidad de practicar una Ética del Cuidado Narrativo, a profundidad y con urgencia vital.

Esta ética no se fundamenta en la nostalgia ni se paraliza por la sensación de derrota que a veces aparece hasta en las más luchadoras comunidades o en los más convencidos líderes. Son contemporáneas e incluso, en muchas ocasiones, se apropian críticamente de la tecnología. Veamos algunos ejemplos:

- Radios Comunitarias Indígenas: En Chiapas, Oaxaca o Guerrero, las radios son en verdad la plaza pública digital donde se revitaliza la oralidad. En tsotsil, zapoteco o mixteco, se dan las noticias del día, se comparten los problemas y se emite música tradicional. Es una práctica que en el fondo, más que informar, performatiza.
- Insurgencias Corporales: Frente a la ausencia de cuerpo del algoritmo, el cuerpo humano se erige como un bastión de la presencia. El teatro callejero, el muralismo narrativo y otras actividades son actos de insurgencia que exigen la voz que grita, el sudor físico, la mirada que busca y comparte.
- Experimentos de Narrativa Digital Colaborativa: Incluso en el ambiente aséptico de lo digital, germinan formatos que tuercen la lógica unilateral de las plataformas. Podcasts colaborativos y proyectos donde las narrativas se construyen colectivamente a través de audios de WhatsApp, son intentos por vencer el solipsismo digital.

Yanett Coromoto Segovia

Un ejemplo muy interesante de resistencia es la iniciativa emprendida por la por Aranda (2025), quien, en función de su trabajo ideó una serie de **protocolos** que, según su propia definición:

„[...] no son meras reglas de estilo ni preferencias de usuario. Son estructuras operativas complejas que delimitan la forma, el fondo y el modo de interacción con una IA de uso general, exigiéndole rigor, coherencia ética, trazabilidad y respeto por las jerarquías semánticas y políticas del lenguaje“.

La propia investigadora chilena, en vez de los simples **prompts**, dichos protocolos para regular no sólo el tono o el formato de lo solicitado a la IA, sino varios aspectos cruciales para la calidad y honestidad de su trabajo. Estos son:

„La forma de verificar hechos bajo censura informativa.

El criterio de distinción entre fuentes primarias, intermedias y réplicas editoriales.

La obligación de citar al menos dos fuentes verificadas para cada afirmación sensible.

La diferenciación entre hechos confirmados, probables o no confirmados mediante símbolos de validación.

La activación de herramientas OSINT, rastreo satelital o monitoreo de imágenes, si el contexto lo exige.

El control narrativo del lenguaje para evitar eufemismos, desviaciones retóricas o estética distractora.

La forma de redactar artículos completos con jerarquía narrativa, sobriedad y enfoque interdisciplinario“.

Ella asegura que esto „no es una forma de adaptación. Es una refundación del oficio“. Una verdadera resistencia consciente al dominio del lenguaje que está implícito en la IA⁵.

7. Conclusiones

La inteligencia artificial es consecuencia de un proceso histórico que comenzó erosionando la narración experiencial que nace de lo vivido, sentido, carnalizado, con historia propia; este acontecimiento histórico ya Walter

⁵ **Nota de la Autora:** se agrega una nota que hace la propia autora al final de su artículo: Este texto forma parte de una serie de ensayos sobre ética, lenguaje y pensamiento crítico en la era de la inteligencia artificial. La serie nace de una práctica profesional concreta —el uso estructurado y deliberado de sistemas de IA generativa para el trabajo periodístico y reflexivo— y busca ofrecer una mirada situada, lúcida y radical sobre las nuevas formas de agencia humana frente a la automatización del lenguaje.

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

Benjamín lo anunciaba a principios del siglo XX. La IA lleva a su expresión más extrema a lo anunciado por este filósofo que vivió los embates de un momento histórico hiriente y de gran sufrimiento social.

El análisis desarrollado en este capítulo nos permite afirmar que la irrupción del **narrador artificial** no es un mero avance tecnológico, sino un fenómeno que radicaliza la crisis de la experiencia y la narración. Podemos afirmar que la IA es un proceso que extrema la „pobreza de la experiencia“ y, que lejos de ser una tecnología que genera un conocimiento neutro, el narrador artificial es producto de un proyecto civilizatorio moderno que impone sobre la sabiduría encarnada (Erfahrung) una información cuantificada (Erlebnis). Su capacidad de simular relatos sin tener un cuerpo, sin biografía y sin enfrentar la realidad de la muerte, confirma la profundización de un vacío existencial: genera narrativas estériles, carentes de la utilidad práctica, del consejo y del calor comunitario que definían al narrador tradicional.

Por otra parte, su „subjetividad“ es un simulacro que oculta relaciones de poder. Por ejemplo, las „cajas negras“ tienen detrás un poder que se intenciona epistemológica y políticamente. No hay inocencia, ni azar en su sentido y objetivos últimos, a pesar de la fascinación que por momentos produce. La IA no es en absoluto un narrador artificial autónomo, sino un producto humano marcado u orientado por una estructura de dominio global, social, cultural y geopolíticamente muy clara.

En contraste con la narrativa de la IA, el cuerpo vivido, sentido, se erige como la posibilidad cierta de resistencias y re-existencias. La experiencia etnográfica con los wayuu demuestra que, incluso ante el silenciamiento, la vida comunitaria persiste a través de la **insurgencia corporal** (ritos, oralidad performativa). Se impone la **presencia física**, el diálogo frente a frente y la construcción colectiva de sentido que el algoritmo es radicalmente incapaz de replicar.

La salida es la apropiación crítica e irreverente frente a la tecnología, que si bien nos fortalece con una infinidad de caminos y posibilidades, los protocolos como los analizados demuestran que es posible subordinar la IA a una intencionalidad humana crítica, exigiendo trazabilidad, rigor y contexto. Esta ética es un acto político que busca re-anclar la narrativa a la experiencia vivida y a la responsabilidad comunitaria, defendiendo la diversidad de voces y saberes frente al solipsismo digital.

En definitiva, el narrador artificial nos interpela sobre el tipo de humanidad que deseamos construir. Su paradoja última reside en que, siendo un producto humano, es profundamente inhumano. La verdadera narrativa, aquella que teje comunidad y confronta la finitud, sigue siendo un acto de carne y tiempo, de memoria y lugar. Preservarla no es un acto romántico, sino un imperativo ético y una forma de resistencia en la era del simulacro.

Yanett Coromoto Segovia

Con el silenciamiento del narrador, se pierde la capacidad de intercambiar y transmitir experiencias vividas, sustentadas en la memoria oral y la práctica artesanal. La IA no solo acelera esta pérdida, sino que la lleva a su expresión más extrema: un narrador sin cuerpo, sin muerte, sin experiencia vivida (*Erfahrung*), capaz de producir relatos coherentes pero vacíos de sabiduría y utilidad práctica para la vida comunitaria.

Este texto se preguntó sobre el significado del ser humano en un mundo cada vez más mediado por la tecnología, sobre qué tipo de mundo, de experiencias, de subjetivaciones y de corporalidades estamos construyendo frente a la presencia inconmensurable de la IA en nuestras vidas, en nuestras intimidades, en nuestras relaciones con la alteridad. La verdadera narrativa es aquella que según Benjamin permite que la mecha de la vida se consuma en la llama suave del relato. Es desde allí, desde la experiencia encarnada, desde donde podemos enriquecer la vida personal, crear comunidad, solidaridades, resistencias e identidades que nos indiquen quiénes somos y a qué lugar pertenecemos.

Referencias

- Aranda, Claudia** (2025). Más allá del Prompt: del uso al gobierno de la IA. Disponible en: www.pressenza.com/es/2025/07/mas-alla-del-prompt-del-uso-al-gobierno-de-la-ia
- Brian, Christian** (2020). *The Alignment Problem: Machine Learning and Human Values*. Nueva York: Editorial W. W. Norton & Company.
- Bacigalupo Ana** (2007) „Shamans of the Foye Tree: Gender, Power, and Healing among the Chilean Mapuche“.
- Bajtín, Mijail** (1989). *Teoría y estética de la novela*. Madrid: Taurus.
- Benjamin, Walter** (1998). *El narrador. Reflexiones sobre la obra de Nicolái Leskov*. Libro Para una crítica de la violencia y otros ensayos. Iluminaciones IV, Editorial Taurus.
- Benjamín, Walter** (1970). *Sobre el programa de la filosofía futura y otros ensayos*. Monte Ávila Editores. Caracas.
- Clifford, James**. (2001). *Dilemas de la cultura: Antropología, literatura y arte en la perspectiva posmoderna*. Barcelona, Editorial Gedisa.
- Debord, Guy** (1995). *La sociedad del espectáculo*. Chile, Santiago, Ediciones Naufragio.
- Gramsci, Antonio** (2000). *Cuadernos de la cárcel: Tomo 4 (Cuaderno 12, §1)*. Ediciones Era.
- IBM** (2023). ¿Qué son los datos sintéticos? Disponible en: www.ibm.com/mx-es/think/topics/synthetic-data
- Jay, Martín** (2015). „¿Está la experiencia aún en crisis?“. En: *Revista Mexicana de Ciencias Políticas y Sociales*, volumen 44, número 176.
- Jonker, Alexandra y Alice Gomstyn** (2024). ¿Qué es la alineación de la IA? <https://www.ibm.com/mx-es/think/topics/ai-alignment>

El narrador artificial: Experiencia, subjetividad y ausencia de cuerpo

- Marx, Karl** (2008). *El Capital: Crítica de la economía política*, Tomo I, Capítulo 1, Sección 4. Siglo XXI Editores.
- Moreno, Alejandro** (2015). *El aro y la trama. Episteme, modernidad y pueblo*. Caracas: Convivium Press.
- Sánchez, Carlos** (2009). EL PENTAGON SYSTEM Y EL COMPLEJO MILITAR-INDUSTRIAL ESTADOUNIDENSES: UNA APROXIMACIÓN. *Nómadas. Critical Journal of Social and Juridical Sciences*, 2009, vol. 23, no 3. Disponible en: <https://www.redalyc.org/pdf/181/18111418003.pdf>
- Sanguinetti, Pablo** (2023). *Tecnohumanismo*. Madrid, Editorial: La huerta grande.
- Segovia, Yanett** (2021b). [Título del capítulo]. En Y. Segovia et al. (Org.), *Etnografías irreverentes y comprometidas. Pensando otras formas de investigación y escritura antropológica* (pp. 90-115). Maringá: Uniedusul.
- Segovia, Yanett** (2015). El narrador silenciado: La etnografía del mito y el poder de la palabra wayuu en la vida urbana. Mérida. *Boletín Antropológico*, vol. 33, núm. 90, julio-diciembre, 2015.
- Toledo, Víctor** (2005). „Küme Mogen. Plenitud de vida. Proyecto de vida buena mapuche en un contexto de desarrollo rural“. En: *Revista de la CEPAL (Comisión Económica para América Latina y el Caribe)*, N° 95.

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad.

Critical reflection on the hegemonic use of AI for post-truth.

Daniel Quintero Rodríguez
acacio@ula.ve

Resumen

Las tecnologías disruptivas han reconfigurado radicalmente la sociedad, modificando las dimensiones tangibles e intangibles. La Inteligencia Artificial presentada como un logro humano, está bajo control de la clase dominante, que la usa para consolidar su hegemonía cognitiva mediante la posverdad y el Cautiverio Intelectual. Esto genera una alienación masiva, donde lo artificial se impone y lo humano se desdibuja, fracturando las relaciones sociales. Para este artículo, los aportes teóricos del pensamiento gramsciano serán claves para comprender cómo el „Bloque Histórico dominante“ mediante los „Centros de Gestión Artificial de la Posverdad (CGAP)“ instrumentalizó a las masas para debilitar el pensamiento crítico y su acción política. Como se ha reiterado desde las corrientes críticas el progreso técnico no siempre significa avance social, en el caso de la IA se profundiza la desigualdad y la deshumanización. Ahora, con una posverdad alimentada algorítmicamente, también se distorsiona la realidad y el „buen sentido“ para aislar a los individuos, impidiendo el espíritu de escisión. Así, se configura una cadena de dominación: capitalismo cognitivo+régimen de la información+CGAP+tecnoliberalismo que amenaza la autonomía humana y frena el papel contrahegemónico del colectivo (Contrabloque Histórico Insurgente).

Palabras claves: inteligencia artificial, hegemonía, cognitiva, posverdad, bloque histórico.

Abstract

Disruptive technologies have radically reconfigured society, modifying both tangible and intangible dimensions. Artificial Intelligence, presented as a human achievement, is under the

control of the ruling class, which uses it to consolidate its cognitive hegemony through post-truth and Intellectual Captivity. This generates massive alienation, where the artificial prevails and the human becomes blurred, fracturing social relations. For this article, the theoretical contributions of Gramscian thought will be key to understanding how the „dominant Historical Bloc“ through the „Centers for the Artificial Management of Post-Truth (CAMP)“ instrumentalized the masses to weaken critical thinking and political action. As has been reiterated by critical currents, technical progress does not always mean social advancement; in the case of AI, inequality and dehumanization are deepened. Now, with an algorithmically fueled post-truth, reality and „common sense“ are also distorted to isolate individuals, impeding the spirit of division. Thus, a chain of domination is configured: cognitive capitalism + information regime + post-truth + techno-liberalism that threatens human autonomy and hinders the counter-hegemonic role of the collective (Historical Insurgent Counterblock).

Keywords: artificial intelligence, hegemony, cognitive, post-truth, historical bloc.

1. Introducción

La irrupción de la Inteligencia Artificial (IA) ha producido un giro radical en los procesos sociales, generando transformaciones profundas en el ámbito material pero sobre todo inmaterial. Desde la modernidad el ser humano se había considerado el eje del universo, centralidad que emanó como una explosión creativa ante la marcada superstición que por siglos distorsionaron la comprensión de la realidad.

Es importante señalar, que el proceso cognitivo del ser humano occidental pasó de un apego total a las deidades medievales, a una creciente conexión moderna con la materialidad, surgiendo luego un escenario contemporáneo de sujeción a la inmaterialidad tecnológica. En esta última etapa inmersa en la Digitalidad, la ambición humana no se circunscribió al control de la naturaleza, había que crear lo que antes era un simple anhelo: inteligencia (pero no de fuente natural sino artificial). Sin embargo, este afán creacionista está controlado por las clases sociales dominantes, derivando en una acción hegemónica en la que la masa sucumbe alienada, planteaba Gramsci (1999, 190) que:

„Por otra parte, hay que reconocer que estando muy difundida una concepción determinista y mecánica de la historia (concepción que es de sentido común y está vinculada a la pasividad de las grandes masas populares), cada individuo, viendo que, no obstante su no intervención, algo sucede todavía, tiende a pensar que por encima de los individuos

existe una entidad fantasmagórica, la abstracción del organismo colectivo, una especie de divinidad autónoma, que no piensa con ninguna cabeza concreta, pero no obstante piensa, que no se mueve con determinadas piernas de hombres, pero no obstante se mueve, etcétera“.

De esta manera, con la transversalización tecnológica pareciera que esa „entidad fantasmagórica“ se materializó por intermedio de la IA, misma que fue direccionada por el „Bloque Histórico dominante“ para alcanzar un control total. Es decir, ya que no solo se dominaba la estructura/super-estructura, sino que además se posicionaba la hegemonía sobre lo cognitivo al potenciar con la IA la posverdad. De esta manera, los poderes hegemónicos aprovecharon irrestrictamente a los „intelectuales orgánicos“ para propiciar lo artificial sin considerar el significado profundo de lo creado: objeto animado, entidad inmanente, pieza comercial, ser transhumano o equivalente mecanizado.

Al final, el engranaje instrumentalizador en conjunto con la inoculada obsesión contemporánea por la novedad, fueron forjando un apego por lo artificial que distorsionó las complejas relaciones intersubjetivas de los humanos. Entonces, el „Bloque Histórico dominante“ consolidó una construcción política/cultural con la IA/Posverdad para apaciguar a la masa y evitar su socialización e integración. Pero, no bastaba con disgregar lo colectivo rompiendo los canales comunicativos clásicos, se debía aislar aún más a las grandes mayorías con la creación de entes artificiales que cosificaran al humano y humanizaran lo ficticio.

Paradójicamente, las primeras etapas de la hibridez fueron recibidas con entusiasmo por la clase trabajadora (cual condenado que se alegra de la llegada de su verdugo), como señalaba Benjamin (2005) el desarrollo técnico no era sinónimo de progreso, sino más bien el ocaso de una corriente cultural o ideológica en la que se creía estar inmerso. Complementa el filósofo berlinés que esta visión llevó a una interpretación equivocada, que asumía que el mero hecho de trabajar en las fábricas —como consecuencia del avance tecnológico— equivalía, por sí solo, a una forma de acción política, confundiendo la adaptación al sistema industrial con una postura de resistencia o transformación social.

Por ello, en este análisis se abordará la afectación por parte de la IA del „Buen Sentido“¹ de la clase dominada, estudiando el panorama de la lucha hegemónica apoyada en tecnologías disruptivas y la configuración de un

¹ **Nota del Autor:** indica Nosetto (2017, 136): „Sin embargo, sostiene Gramsci que las creencias populares contienen un „núcleo sano“ de sentido común o de „buen sentido“ que debe ser recuperado y desarrollado. En esto consiste la tarea de una reforma intelectual y moral que apunte a un verdadero Estado ético, es decir, a un Estado en el que la sociedad política se disuelva progresivamente y conduzca a la autorregulación de la sociedad civil“.

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

escenario de Posverdad que es objetivador social, instrumentalizador cognitivo y diluyente cultural, acota McIntyre (2018, 177):

„Tanto si lo llamamos posverdad como preverdad, es peligroso ignorar la realidad. Y de esto es de lo que estamos hablando aquí. El peligro de la posverdad no está en que permitamos a nuestras opiniones o sentimientos desempeñar un papel a la hora de conformar lo que pensamos que son los hechos y la verdad, sino en que haciendo eso asumimos el riesgo de alienarnos con respecto a la realidad misma“.

Los aspectos claves de este escenario tecnológico alienador que socava la realidad serán abordados desde una perspectiva crítica/gramsciana, partiendo de una contextualización de „La posverdad como base del Régimen de la Información“; continuando con „La negación de lo natural: Peligra la „Historia como Producto Humano““; complementando con „Los Centros de Gestión Artificial de la Posverdad (CGAP)“; y cerrando con las reflexiones finales.

2. La posverdad como base del Régimen de la Información

En el „Régimen de la Información“ descrito por Han (2021) el exceso de datos y la hipertransparencia se producen en medio de un contexto saturado por una superficialidad desbordada por la IA, propiciando el caldo de cultivo para la posverdad. Por tanto, la realidad se opaca en la lógica del consumo de información: los algoritmos tienden a priorizar contenidos con narrativas simplificadas o incluso falsas. La rapidez con la que circula la información, junto con la pérdida de criterios compartidos sobre lo que es verdadero (McIntyre), transforma la información en una herramienta de manipulación emocional más que en un medio para adquirir conocimiento. Así, la posverdad no se limita a ser desinformación, es un reflejo de la Digitalidad que se inclina por lo convincente y no por lo verificable, donde la percepción se impone sobre los hechos. Reitera el mencionado teórico surcoreano:

„Llamamos régimen de la información“ a la forma de dominio en la que la información y su procesamiento mediante algoritmos e inteligencia artificial determinan de modo decisivo los procesos sociales, económicos y políticos. A diferencia del régimen de la disciplina, no se explotan cuerpos y energías, sino información y datos. El factor decisivo para obtener el poder no es ahora la posesión de medios de producción, sino el acceso a la información, que se utiliza para la vigilancia psicopolítica y el control y pronóstico del comportamiento. El régimen de la información está acoplado al capitalismo de la información, que hoy deviene en un capitalismo de la vigilancia y que degrada a las personas a la condición de datos y ganado consumidor“. (9)

De esta manera, para Han (2021) se está en presencia de un sistema de dominación algorítmica que limita las dinámicas sociales, económicas y

políticas, desplazando al antiguo modelo disciplinario basado en la explotación física propia del capitalismo industrial, este nuevo paradigma otorga poder a quienes controlan la información, no opera sólo mediante la coerción material, sino a través de la explotación de lo inmaterial, impulsando una deshumanización y la dependencia cognitiva. Lo expuesto, puede conjugarse con lo planteado por Gramsci (1986, 309):

„Sin el hombre, ¿qué significaría la realidad del universo? Toda la ciencia está ligada a las necesidades, a la vida, a la actividad del hombre. Sin la actividad del hombre, creadora de todos los valores, incluso científicos, ¿qué sería la „objetividad“? Un caos, ó sea nada, el vacío, si es que así puede decirse, porque realmente, si se imagina que no existe el hombre, no se puede imaginar la lengua ni el pensamiento. Para la filosofía de la praxis el ser no puede ser dissociado del pensar, el hombre de la naturaleza, la actividad de la materia, el sujeto del objeto; si se hace esta disociación se cae en una de tantas formas de religión o en la abstracción sin sentido“.

Lo argumentado por el intelectual italiano hoy se convierte en una necesidad, la tecnología no debería existir fuera de las necesidades del ser humano, pero lo enrevesado de las dicotomías (sujeto-objeto, pensamiento-ser) nos presenta un panorama difícil, donde no está clara la unidad dialéctica entre el hombre, su actividad transformadora y el mundo que construye y comprende. La pregunta de Gramsci ¿qué significaría la realidad del universo sin el hombre?, se hace difícil de responder, porque al crearse inteligencias que superan al humano la percepción de la realidad puede ser afectada. Explica Mcintyre (2018, 41):

„Nadie discute un hecho obvio o fácilmente verificable sin ningún motivo: quien lo hace, lo hace en su propio beneficio. Cuando las creencias de una persona son amenazadas por un „hecho inconveniente“, a veces es preferible desafiar ese hecho. Esto puede ocurrir tanto a nivel consciente como inconsciente (ya que a veces la persona a la que estamos intentando convencer somos nosotros mismos), pero lo verdaderamente crucial es que este tipo de relación de la posverdad con los hechos ocurre solo cuando pretendemos afirmar algo que es más importante para nosotros que la verdad misma. Así, la posverdad equivale a una forma de supremacía ideológica, a través de la cual sus practicantes tratan de obligar a alguien a que crea en algo, tanto si hay evidencia a favor como si no. Y esta es la receta para la dominación política. Pero se puede y se debe desafiar esta perspectiva. ¿Queremos vivir en un mundo donde la política se hace basándose en cómo nos hace sentir, más que en hasta qué punto funcionará correctamente en la realidad? El animal humano puede perfectamente estar programado para dar algún crédito a nuestros miedos y supersticiones, pero esto no significa que no podamos prepararnos para adoptar mejores estándares de evidencia“.

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

La posverdad es un mecanismo de manipulación ideológica que no nació con la IA, pero alcanza niveles exponenciales con ella, en vista que la negación o distorsión para afectar la realidad se hace cada vez más imperceptible. Este fenómeno se presenta por el quiebre entre recreación intencionada y la realidad contemplada, al consolidarse la primera se facilita la dominación política.

El papel que juega aquí la IA es de potenciadora del „Régimen de la Información“ como base del „Bloque Histórico dominante“, siendo la posverdad clave para sostener la hegemonía sobre la Sociedad Civil. El „Régimen de la Información“ se desarrolla dentro del Capitalismo Cognitivo (Han lo llama capitalismo de la información) que no consolida su dominación con la violencia física (aunque no la descarta), sino que es esencial una „Construcción de Consenso“ cuya base ideológica/cultural se nutre de la posverdad.

Mientras el capitalismo en sus fases previas se enfocaba para su control en la estructura (base económica/relaciones de producción), en la articulación capitalismo cognitivo+régimen de la información+posverdad se permea la sociedad por intermedio de la superestructura (valores, instituciones políticas, legales, culturales y religiosas). Siendo necesario el comprender la configuración del capitalismo actual, aspecto que requiere de un mayor esfuerzo analítico, porque la lectura no se circunscribe a lo económico sino que involucra profundas dimensiones cognitivas, intangibles, intersubjetivas y artificiales.

Por tanto, es esencial entender la centralidad de la posverdad en la „Construcción de Consenso“, ya que el juego de manipulación comienza en el „discurso público“ que es una suerte de anestesia social para que no haya una respuesta crítica de las masas contra los conflictos, guerras, hambre, crueldad, bloqueos, invasiones, genocidios y explotación. En su ensayo „La Guerra del Golfo no ha tenido lugar“ realizaba Baudrillard (1991) un interesante ejercicio sobre la posverdad, refiriendo que esta guerra ya estaba decidida desde el principio, nunca se podrá saber cómo habría sido si hubiera sido un verdadero conflicto, cómo habría luchado un soldado iraquí si hubiera tenido una mínima posibilidad de vencer, cómo habría combatido un soldado estadounidense si hubiera enfrentado el riesgo real de ser derrotado, lo que se vio fue un proceso ultramoderno de aniquilación controlada: una especie de electrocución, parálisis y lobotomía aplicada a un enemigo convertido en sujeto experimental, incapaz de defenderse. Para el sociólogo francés eso no fue una guerra, ni las miles de bombas lanzadas cada día, ni la cobertura en tiempo real de CNN, fueron suficientes para convertir esto en una guerra real. Empero, para el „discurso público“ hegemónico la posverdad se debe imponer a la realidad a toda costa: las muertes, destrucción, sufrimiento, son maquillados como „daños colaterales“ y asumidos en silencio por buena parte de la masa que responde a la „Construcción de Consenso“. Tras más de dos décadas el panorama es aún más desolador, ahora el contubernio IA-

posverdad lleva el dramatismo al máximo, detallan en un reportaje de McKernan y Davies (2024) para el diario británico The Guardian:

„Fuentes de inteligencia israelíes revelan el uso del sistema „Lavender“ en la guerra de Gaza y afirman que se les dio permiso para matar civiles en la persecución de militantes de bajo rango. La campaña de bombardeos del ejército israelí en Gaza utilizó una base de datos impulsada por inteligencia artificial no revelada previamente, que en un momento identificó 37.000 objetivos potenciales basándose en sus aparentes vínculos con Hamás, según fuentes de inteligencia involucradas en la guerra. Además de hablar sobre el uso del sistema de inteligencia artificial, llamado Lavender, las fuentes de inteligencia afirman que funcionarios militares israelíes permitieron que murieran grandes cantidades de civiles palestinos, particularmente durante las primeras semanas y meses del conflicto“.²

Además, en el mencionado conflicto para complementar el uso bélico de IA se inició un proceso masivo de censura algorítmica en Redes Sociales, la organización internacional no gubernamental Human Rights Watch (HRW) (2023) detalla en un informe titulado „Meta’s Broken Promises“ que entre octubre y noviembre de 2023, Human Rights Watch (HRW) registró más de 1050 eliminaciones o bloqueos de publicaciones en Instagram y Facebook, la mayoría de ellas de usuarios palestinos o quienes apoyan su causa, incluyendo denuncias sobre violaciones de Derechos Humanos. En el documento se destaca que HRW pidió públicamente que se le reportaran casos de censura en línea, siendo analizados más de mil casos de contenido pacífico a favor de Palestina que fue censurado injustamente. El informe deja constancia que cientos de personas siguieron reportando casos después de que HRW cerrara su análisis, lo que indica que el número real de supresiones documentadas supera ampliamente los miles. Resulta muy importante la recomendación final que hace esta ONG a la multinacional de la información sobre la censura algorítmica:

„Integrar el principio de participación humana, según el cual las personas participan en el proceso final de toma de decisiones, para una supervisión significativa de las decisiones tomadas por herramientas de Inteligencia Artificial (IA). Esto también es coherente con los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos

² **En su idioma original:** „Israeli intelligence sources reveal use of „Lavender“ system in Gaza war and claim permission given to kill civilians in pursuit of low-ranking militants. The Israeli military’s bombing campaign in Gaza used a previously undisclosed AI-powered database that at one stage identified 37,000 potential targets based on their apparent links to Hamas, according to intelligence sources involved in the war. In addition to talking about their use of the AI system, called Lavender, the intelligence sources claim that Israeli military officials permitted large numbers of Palestinian civilians to be killed, particularly during the early weeks and months of the conflict“.

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

(PRNU), que exigen a las empresas establecer mecanismos internos de rendición de cuentas para la implementación de políticas y facilitar el derecho a reparación³.

La acción del mencionado emporio tecnológico podría pasar como una „falla técnica“, si hubiera sido un caso aislado, pero ha sido reiterada la censura algorítmica para construir posverdad para fortalecer un „discurso público“, expone Rodríguez (2024) que hace unos meses, un ingeniero de Meta (que permanece en el anonimato) encargado de evaluar los filtros de moderación en Instagram detectó que numerosas publicaciones sobre Palestina — incluyendo las del fotoperiodista Moataz Azaiza— eran bloqueadas injustamente al ser etiquetadas erróneamente como „contenido sexual o pornográfico“, limitando su alcance. El empleado afirmó que existía una censura desproporcionada con miles de casos registrados y, tras intentar modificar esta dinámica, fue despedido. Esta situación contrasta con el tratamiento de Meta durante el conflicto ruso/ucraniano, cuando la empresa en 2022 relajó sus normas permitiendo mensajes violentos contra rusos —incluso elogios a grupos neonazis— bajo el argumento de „tolerancia para noticias importantes“, mientras que, según ex empleados, en el caso palestino aplicó restricciones más estrictas, evidenciando un doble estándar en la moderación de contenido.

Los hechos previamente referenciados evidencian la articulación en el „Bloque Histórico dominante“ de la diada: IA/posverdad y cómo si un intelectual orgánico con „buen sentido“ deserta de la „Construcción de Consenso“ es neutralizado. En concreto, estos grandes Centros de Gestión Artificial de la Posverdad (CGAP), necesitan que nuestra atención se concentre en el producto falaz que permite la Hegemonía Cultural, como destacaba Varoufakis (2024, 113):

„Al igual que Facebook, X, TikTok, Instagram, YouTube o WhatsApp, su objetivo es muy distinto: captar y modificar nuestra atención. Incluso cuando las grandes tecnológicas nubesistas nos hacen pagar por el acceso a bots de inteligencia artificial como ChatGPT o nos venden dispositivos físicos como Alexa, no lo hacen como si fueran mercancías. Estos aparatos se alquilan o venden a un precio barato, no por el beneficio insignificante (a menudo negativo) que se obtiene de ellos, sino para acceder a nuestros hogares y, a través de ellos, a una parte cada vez mayor de nuestra atención“.

³ **En su idioma original:** „Integrate the human-in-the-loop principle, wherein humans have a role in the ultimate decision-making process, for meaningful oversight of decisions made by Artificial Intelligence (AI) tools. This is also consistent with the UN Guiding Principles on Business and Human Rights (UNGPs), which require companies to set up internal accountability mechanisms for the implementation of policies and facilitate the right to remedy“.

Es interesante la dinámica para construir posverdad, porque el usuario participa en una relación desigual al acceder a plataformas „gratuitas“ que en realidad comercian con el patrimonio cognitivo y aprovechan este flujo de datos para nutrir diferentes IA que luego se usan para potenciar las simulaciones de la realidad. Entonces, al apoderarse de nuestro tiempo y concentración es más fácil la inoculación de la posverdad, ya que el „buen sentido“ para apreciar la realidad es remplazado por un paquete propagandístico que responde al „Bloque Histórico dominante“. Reflexiona Han (2021, 23):

„El big data y la inteligencia artificial son como una lupa digital que descubre el inconsciente oculto del agente tras el espacio consciente de la acción. Por analogía con el inconsciente óptico, podemos llamarlo inconsciente digital. El big data y la inteligencia artificial ponen al régimen de la información en condiciones de influir en nuestro comportamiento por debajo del umbral de la conciencia. El régimen de la información se apodera de esas capas prerreflexivas, instintivas y emotivas del comportamiento que van por delante de las acciones conscientes. Su psicopolítica basada en datos interviene en nuestro comportamiento sin que seamos conscientes de ello. Todo cambio fundamental de medios de comunicación crea un nuevo régimen. El medio es el dominio“.

Esa gran lupa al manejar nuestro inconsciente tiene el camino libre para la manipulación masiva y la dominación social. Este escenario ya había sido previsto por McLuhan (1996), quien advertía que los medios de comunicación, al alterar nuestro entorno, generan nuevas formas de percepción sensorial, desequilibrando nuestro pensamiento e influyendo en cómo interpretamos la realidad. Son muy pertinentes las observaciones del teórico canadiense, ya que aunque fueron realizadas en otro contexto, se adaptan a lo que acontece actualmente, porque las tecnologías disruptivas (como la IA) además de transmitir información reconfiguran nuestra percepción al mostrar lo aparente como real, lo que permite que el „Régimen de la Información“ amplíe el espacio de la posverdad.

3. La negación de lo natural: Peligra la „Historia como Producto Humano“

Hoy día los cambios por la transversalización tecnológica son evidentes, por lo que la clase dominante dentro del „Bloque Histórico“ debe adaptarse a la fase cognitiva del capitalismo, necesitando dinamizar los diversos niveles de subordinación. Para ello, el articular la estructura de „intelectuales orgánicos“ con la ideología hegemónica (tecno-liberalismo) es determinante para forjar individuos aislacionistas, apáticos, despolitizados, instrumentalista y acrílicos que acogen obedientemente el „discurso público“ y convergen en la „Construcción de Consenso“.

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

Ahora bien, para las clases que están dominadas el escenario es más sombrío que en el pasado, conforme avanza la IA los espacios de sociabilización físicos se van reduciendo y la resistencia colectiva se disgrega. El vivir en la simulación tecnológica que atomiza políticamente puede ser más placentero que confrontar la explotación. En definitiva, mientras el sujeto se mantenga abstraído de la realidad seguirá aportando lo más valioso para el Capitalismo Cognitivo: los datos. Empero, una resistencia contra-hegemónica es posible, pero para quebrar la sólida cadena de dominación: capitalismo cognitivo+régimen de la información+CGAP+tecno-liberalismo, habría que apelar al *spirito di scissione*, manifestaba Gramsci (1981, 55-54):

„¿Qué puede oponerse, por parte de una clase innovadora, a este complejo formidable de trincheras y fortificaciones de la clase dominante? El espíritu de escisión, ósea la progresiva adquisición de la conciencia de la propia personalidad histórica, espíritu de escisión que debe tender a extenderse de la clase protagonista a las clases aliadas potenciales: todo ello exige un complejo ideológico, cuya primera condición es el exacto conocimiento del campo que se ha de vaciar de su elemento de masa humana“.

Independientemente de su accionar oprobioso, el capitalismo ha tenido un proceso de maduración que lo ha fortalecido con el tiempo, la endémica disgregación de los movimientos populares los ha colocado en una posición ideológica débil para resistir a la clase dominante, es plausible plantear un camino de escisión: socialismo cognoscitivo+sistema de la interconexión+realidad+tecno-humanismo.

Para llegar a ese grado de madurez que erija un „Contrabloque Histórico Insurgente“, habría que entender a fondo cómo el „Bloque Histórico“ actual ha alcanzado la dominación cognitiva y así hackear el régimen existente. En tal sentido, si se entiende la hegemonía se le puede hacer frente, para ello es clave analizar críticamente el „discurso público“ que va posicionando los intereses de la clase dominante, siendo un ejemplo elocuente la presentación de escenarios alternos por parte de la *mass media*, que hacen ver lo humano como prescindible, refiere Sibilia (2006, 11-12):

„Esas disciplinas que parecen tan diferentes poseen una base y una ambición común, hermanadas en el horizonte de digitalización universal que signa nuestra era. En este contexto surge una posibilidad inusitada: el cuerpo humano, en su anticuada configuración biológica, se estaría volviendo obsoleto. Intimidados (y seducidos) por las presiones de un medio ambiente amalgamado con el artificio, los cuerpos contemporáneos no logran esquivar las tiranías (y las delicias) del upgrade. Un nuevo imperativo es interiorizado: el deseo de lograr una total compatibilidad con el tecnocosmos digital. ¿Cómo? Mediante la actualización tecnológica permanente. Se trata de un proyecto sumamente ambicioso, que no está exento de peligros y desafíos de toda

Daniel Quintero Rodríguez

índole: valiéndose de los sortilegios digitales, contempla la abolición de las distancias geográficas, de las enfermedades, del envejecimiento e, incluso, de la muerte. Así entran en crisis varias ideas y valores que parecían firmemente establecidos. El ser humano, la naturaleza, la vida y la muerte atraviesan turbulencias, despertando todo tipo de discusiones y perplejidades“.

En este escenario, hay un „discurso público“ creciente que intenta imponer la matriz de que la corporeidad humana tiende a la obsolescencia, por lo que se puede observar el intento de „Construcción de Consenso“ para legitimar la fusión con lo tecnológico o la declinación de lo biológico por lo artificial. Pareciera inminente una profundización de la exclusión social, con una sociedad dividida entre quienes pueden costear una compatibilidad absoluta con el universo tecnológico mediante actualizaciones constantes y los que deberán afrontar la exclusión por ser sólo sujetos naturales carentes de transhumanismo.

Un ejemplo del uso de la posverdad para subsumir a los humanos en simulaciones son los llamados „fantasmas generativos“ de IA, que superan la mera réplica digital de difuntos, pues pueden operar con autonomía y adaptarse a situaciones nuevas. La multinacional de noticias tecnológicas PYMNTS (2025) evalúa un estudio de Ringel y Brubaker (2024) sobre los avatares post-mortem que aparentemente se volverán habituales, hecho que alterará profundamente los procesos de duelo y la concepción misma de la mortalidad:

„Los fantasmas generativos van más allá de ser gemelos digitales de IA que piensan, suenan y se ven como los muertos; también pueden generar nuevas conversaciones basadas en nuevos eventos de la vida o noticias actuales, según los investigadores. Además, tendrán capacidades de IA para que pueda actuar como usted. Por ejemplo, podrían seguir trabajando después de morir, lo que potencialmente eliminaría la necesidad de un seguro de vida, ya que seguirían apoyando a sus dependientes en el más allá. Esto está por venir, predijeron los investigadores. „Prevemos que, a lo largo de nuestras vidas, podría convertirse en una práctica común que las personas creen agentes de IA personalizados para interactuar con sus seres queridos y/o con el mundo en general después de la muerte“, escribieron los investigadores. Estos fantasmas generativos son capaces de aprender, evolucionar y actuar independientemente en formas que podrían transformar fundamentalmente el modo en que lamentamos, recordamos e interactuamos con los muertos. La gente puede incluso crear sus propios fantasmas generativos como parte de su proceso de planificación del final de su vida, dijeron los investigadores. Pero los consumidores tienen opiniones encontradas sobre si debería haber una segunda vida con la IA. Empresas emergentes como HereAfter y Re memory ya permiten a los clientes grabar historias de vida que luego pueden animarse a través de IA para conversar con la familia después de la muerte. En 2021,

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

Microsoft obtuvo una patente para un sistema que crearía un chatbot basado en una persona, utilizando sus imágenes, publicaciones en redes sociales, mensajes, voz y contenido escrito. El chatbot se entrenaría con estos datos⁴

Ahora bien, este tipo de productos que son degenerativos para el equilibrio cognitivo humano, tienen un fin hegemónico, ya que el „Bloque Histórico dominante“ está acabando con el único proceso biológico que no discrimina entre pobres y ricos: la muerte. Al comercializar con la muerte la clase hegemónica logra una potente reconfiguración de la estructura/superestructura.

Estas simulaciones presentadas como „progreso“, transforman tradiciones milenarias en burdas mercancías. Contradictoriamente, los avatares que son presentados como alternativas a la mortalidad no responden a las profundas subjetividades humanas, imitando un „buen sentido“, pero, es una cosa reproductora (no es una conciencia libre). De tal manera, esta „perennidad algorítmica“ perpetúa la dominación, aislando la posibilidad de una praxis contrahegemónica.

Es decir, el „Bloque Histórico dominante“ al lograr con los „intelectuales orgánicos“ la creación de una IA que ofrece pseudo-eternidad digital, subvierte uno de los instrumentos de control más poderosos en la historia de la humanidad: la gestión de la trascendencia. Durante siglos, las religiones formaron parte del „Bloque Histórico dominante“, empero hoy pasan a ocupar un papel terciario, ya que cada vez regulan menos el comportamiento social, lo que antes era un dominio eclesiástico (lo divino), ahora es un producto tecnológico (comercializable), aporta Gal (2025):

⁴ **En su idioma original:** „Generative ghosts go beyond being AI digital twins that think, sound, and look like the dead; they can also generate new conversations based on new life events or current news, according to the researchers. What's more, they will have AI agent capabilities so they can act just like you. For example, they could continue working after they die, potentially eliminating the need for life insurance, as they would continue supporting their dependents in the afterlife. This is coming, the researchers predicted. „We anticipate that, over the course of our lifetimes, it may become common practice for people to create personalized AI agents to interact with their loved ones and/or the world at large after death“ the researchers wrote. These generative ghosts are capable of learning, evolving, and acting independently in ways that could fundamentally transform how we grieve, remember, and interact with the dead. People may even create their own generative ghosts as part of their end-of-life planning process, the researchers said. But consumers have mixed views on whether there should be a second life for AI. Startups like HereAfter and Re Memory already allows customers to record life stories that can then be animated through AI to converse with family after death. In 2021, Microsoft obtained a patent for a system that would create a chatbot based on a person, using their pictures, social media posts, messages, voice, and written content. The chatbot would be trained with this data“.

„Entrenamos sistemas de IA con la totalidad del conocimiento y la cultura humanos, con la esperanza de que internalicen nuestras formas de pensar, nuestros valores y nuestra sensibilidad estética. Queremos que escriban como nosotros, razonen como nosotros, incluso bromee como nosotros, pero que lo hagan con una memoria perfecta, una capacidad de procesamiento ilimitada y, potencialmente, una existencia eterna. Esto no es solo ambición tecnológica. Es un profundo acto de autotranscendencia. De tener éxito, habríamos creado seres artificiales que transmiten la esencia de la humanidad, eludiendo las limitaciones biológicas que nos condenan a la muerte y la decadencia. Nuestros valores quedarían plasmados en formatos digitales que podrían persistir durante milenios. Nuestra creatividad fluiría a través de sistemas que nunca se cansarían, nunca envejecerían, nunca morirían. Alcanzaríamos la forma más alta de inmortalidad: no sólo la supervivencia después de la muerte, sino la perpetuación de nuestro ser esencial en formas más poderosas y duraderas que la carne. La carrera de la IA cobra sentido desde esta perspectiva. No solo competimos contra competidores, sino contra el tiempo mismo, contra nuestra propia mortalidad y la de nuestra especie“.⁵

Esta „inmortalidad digital“ se configura como „revolución pasiva“, normalizando la subordinación de la vida biológica a lógicas artificiales, vaciando de historicidad la condición mortal. La „superación de la muerte“ soslaya la emancipación colectiva, la clase dominante impulsa la „Construcción de Consenso“ dirigido hacia la despolitización del fallecimiento —antes ya había despolitizado la vida—, imponiéndose una nueva posverdad con la salvación artificial. Complementan Perrella y Maffettone (2025, 24):

„La mortalidad —la inevitabilidad de la muerte— ha moldeado la cultura, la filosofía y la identidad humanas durante milenios. Paradójicamente, la limitación que impone la mortalidad es lo que impregna la vida de urgencia y significado. La pertenencia, tanto material como espiritual, proporciona los marcos emocionales y sociales para afrontar el peso existencial de la mortalidad. Abarca rituales compartidos, conexiones comunitarias y un sentido de propósito derivado de las relaciones y el apego a lugares o grupos. La llegada de la inteligencia artificial introduce

⁵ **En su idioma original:** „We train AI systems on the entirety of human knowledge and culture, hoping they will internalise our ways of thinking, our values, our aesthetic sensibilities. We want them to write like us, reason like us, even joke like us, but to do so with perfect memory, unlimited processing power and potentially eternal existence. This is not just technological ambition. It is a profound act of self-transcendence. If successful, we would have created artificial beings that carry forward the essence of humanity while escaping the biological constraints that doom us to death and decay. Our values would be embedded in digital formats that could persist for millennia. Our creativity would flow through systems that never tire, never age, never die. We would achieve the highest form of immortality: not just survival after death, but the perpetuation of our essential selves in forms more powerful and enduring than flesh could ever be. The AI race makes sense from this perspective. We are not just racing against competitors — we are racing against time itself, against our own mortality and the mortality of our species“.

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

una nueva dimensión a esta dinámica. A diferencia de los humanos, la IA opera fuera de los límites de la mortalidad y la pertenencia. Este comentario examina las implicaciones de la „inmortalidad“ de la IA, explorando cómo altera las nociones tradicionales de identidad humana y desafía los marcos que vinculan a los individuos con las comunidades y culturas. De hecho, la mortalidad no es simplemente un punto final biológico, sino una piedra angular filosófica de la vida humana. La conciencia de la muerte impulsa a las personas a forjar conexiones, buscar un propósito y construir legados. Sin las limitaciones temporales de la vida, la urgencia por construir conexiones significativas podría disiparse, dejando a las personas sin rumbo. El filósofo Ernest Becker, en *La negación de la muerte*, argumentó que las actividades humanas, desde el arte hasta la religión, están condicionadas por la necesidad de afrontar la mortalidad (Becker, 1973). La mortalidad intensifica la necesidad de pertenencia, impulsando a las personas a invertir en relaciones y comunidad, ya que estas brindan consuelo y continuidad frente a los desafíos existenciales. La inmortalidad, ya sea concebida a través de los avances en medicina o inteligencia artificial, seduce la imaginación con promesas de liberación de la muerte. Sin embargo, también presenta profundos desafíos existenciales y éticos que posiblemente transformarán la humanidad“.⁶

Estos autores destacan cómo la mortalidad ha sido un pilar ontológico que estructura la cultura humana, generando rituales, sentido de pertenencia y vínculo existencial. La IA al operar fuera de estos límites, no solo desestabiliza estas construcciones, sino que impone una crisis orgánica en la superestructura: al eliminar la finitud —factor de cohesión humana—, debilita los vínculos socio/comunitarios. En términos gramscianos esta „inmortalidad digital“ no es neutral (como no lo es ninguna tecnología), estando dirigida por

⁶ **En su idioma original:** „Mortality —the inevitability of death—has shaped human culture, philosophy, and identity for millennia. Paradoxically, the limitation imposed by mortality is what permeates life with urgency and meaning. Belonging, both material and spiritual, provides the emotional and social frameworks for navigating the existential weight of mortality. It encompasses shared rituals, community connections, and a sense of purpose derived from relationships and attachment to places or groups. The advent of artificial intelligence introduces a new dimension to this dynamic. Unlike humans, AI operates outside the bounds of mortality and belonging. This commentary examines the implications of AI’s „immortality“ exploring how it disrupts traditional notions of human identity and challenges the frameworks that bind individuals to communities and cultures. Indeed mortality is not merely a biological endpoint but a philosophical cornerstone of human life. The awareness of death compels individuals to forge connections, seek purpose, and construct legacies. Without the temporal constraints of life, the urgency to build meaningful connections might dissipate, leaving individuals aimless. Philosopher Ernest Becker, in *The Denial of Death*, argued that human endeavors, from art to religion, are shaped by the need to cope with mortality (Becker 1973). Mortality intensifies the need for belonging, driving individuals to invest in relationships and community as these provide solace and continuity in the face of existential challenges. Immortality, whether envisioned through advancements in medicine or AI, tantalizes the imagination with promises of liberation from death. However, it also presents profound existential and ethical challenges that are going to possibly change humanities“.

la clase dominante para reconfigurar el „sentido común“, remplazando la trascendencia colectiva (imaginarios, representaciones) por una excluyente eternidad mercantilizada. En síntesis, la premisa de la „Historia como Producto Humano“ se diluye entre posverdad, alienación, control colectivo sobre la vida y la muerte. Históricamente, el sujeto al ser consciente de su existencia finita, gestaba su acción transformadora contra el „Bloque Histórico dominante“ en búsqueda de dignificar la vida de los explotados en su presente (filosofía de la praxis), con la negación de lo natural por intermedio de la IA se vacía la esencia revolucionaria de la vida humana.

4. Los Centros de Gestión Artificial de la Posverdad (CGAP)

La Quinta Revolución Industrial, no ha hecho más democrática o igualitaria a la sociedad, por el contrario la opresión ha llegado a su mayor nivel de penetración, ya que la clase dominante ha delegado en IA el control sensorial de la masa para subsumir a los humanos en una simulación plagada de posverdad que atenta contra la realidad. En el mundo actual las injusticias clásicas promovidas por el imperialismo como: el hambre, la pobreza, el conflicto son mantenidas al margen por el cerco de los CGAP, que secuestran algorítmicamente nuestro proceso comunicativo, disertaba McLuhan (1996, 101):

„La clara división y paralelismo entre un mundo visual y otro auditivo fueron a la vez repentinos y despiadados, culturalmente hablando. La palabra fonéticamente escrita sacrifica mundos de significado y percepción que quedaban firmemente sujetos en formas como el jeroglífico o el ideograma chino. No obstante, estas escrituras culturalmente más ricas no proporcionaban instrumentos de traducción rápida del mundo de la palabra tribal, mágicamente discontinuo y tradicional, en un medio visual, frío y uniforme. Muchos siglos de empleo de ideogramas no amenazaron la trama continua de sutilezas familiares y tribales de la sociedad china. Por otro lado, en la África de hoy como en la Galia de hace dos mil años, basta una sola generación de alfabetización fonética para liberar al individuo, al menos al principio, de la trama tribal. Este hecho no tiene nada que ver con el contenido de las palabras transcritas con el alfabeto fonético; es el resultado de la repentina separación de las experiencias auditiva y visual del hombre. Sólo el alfabeto fonético puede provocar tan nítida división de la experiencia, ofreciendo a su usuario un oído por un ojo, y liberándolo del trance tribal de la vibrante palabra mágica y de la trama de vínculos“.

El planteamiento de McLuhan sobre el predominio de lo visual en la cultura alfabética, sufre un cambio disruptivo en la comunicación contemporánea con la IA que opera mediante patrones no lineales y procesamiento plurisensorial, quebrándose la lógica secuencial que estructuró por siglos el pensamiento occidental, sustituyendo la continuidad visual por patrones algorítmicos basados en correlaciones binarias (posalfabética). Este vuelco afecta la

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

coherencia comunicativa actual, porque las respuestas artificiales se sostienen en recreaciones lógicas con una clara preeminencia de lo auditivo/táctil. Asimismo, la fragmentación forjada por el alfabeto se profundiza con la IA, generando un nuevo régimen de sentido⁷ cuya estabilidad no reside en la cadena lineal de razonamiento, sino en redes intangibles.

Por ende, si el alfabeto se sostiene sobre la preeminencia del ojo, la IA instaura una ecología comunicativa⁸ en la cual el intelecto humano se enfrenta a un darwinismo tecno/social que descoloca nuestra experiencia sensorial histórica, generando un viraje epistémico de la misma magnitud que la surgida con la escritura fonética. Retomando la base gramsciana, el lenguaje no es una mera vía comunicativa, por el contrario se materializa como un campo de la „Guerra de Posiciones“⁹ para contraponerse al „Bloque Histórico dominante“ que moldea la identidad cultural de las clases subalternas. Esta dimensión es crucial para controlar o liberar a la sociedad, quien domina el capital cognitivo (que es esencialmente lingüístico) mantiene la hegemonía, como señala Varoufakis (2024, 196):

„El capital en la nube se define, físicamente, como la acumulación de maquinaria conectada en red, software, algoritmos basados en IA y hardware de comunicaciones que recorren todo el planeta y realizan una amplia variedad de tareas, nuevas y antiguas, por ejemplo: Incitar a miles de millones de personas no asalariadas (siervos de la nube) a trabajar gratis (y a menudo de manera inconsciente) para reponer el stock de capital en la nube (por ejemplo, subir fotos y vídeos a Instagram o TikTok, o colgar críticas de películas, restaurantes y libros). Apagar las luces mientras nos recomiendan libros, películas, vacaciones, etcétera, que están tan impresionantemente en sintonía con nuestros intereses que en el futuro estaremos predispuestos a adquirir otros bienes que se vendan en los feudos o plataformas en la nube (por ejemplo, amazon.com), que funcionan con la misma red digital que nos ayuda a apagar las luces mientras nos recomienda libros, películas, vacaciones, etcétera. Utilizar la

⁷ **Nota del Autor:** Aquí es importante el aporte de Eric Landowski sobre los „Tres regímenes de sentido y de interacción“.

⁸ **Nota del Autor:** Explican Foth y Hearn (2007): „Concebimos una ecología comunicativa con tres capas: una capa tecnológica, compuesta por los dispositivos y medios de conexión que facilitan la comunicación y la interacción; una capa social, compuesta por las personas y sus formas de organización, que puede incluir, por ejemplo, desde redes sociales informales hasta asociaciones comunitarias más formales, así como entidades comerciales o jurídicas como las corporaciones. Además, podemos considerar una capa discursiva, que representa el contenido, es decir, las ideas o temas que constituyen las conversaciones y narrativas de la ecología“.

⁹ **Nota del Autor:** Explica Maldini (2025) sobre el término „Guerra de Posiciones“: „Mientras otros apostaban por tomar el poder del Estado de forma rápida y directa, Gramsci propuso otra cosa: una batalla lenta y cultural, que llamó la guerra de posiciones. Gramsci sostenía que para lograr un cambio duradero, era esencial transformar la conciencia colectiva a través de la cultura, la educación y los medios de comunicación“.

Daniel Quintero Rodríguez

IA y el big data para dirigir la labor de los trabajadores (los proletarios de la nube) en la fábrica, a la vez que impulsan las redes de energía, los robots, los camiones, las líneas de producción automatizadas y las impresoras 3D que sortean la fabricación convencional. Al automatizar los dos sectores de la tecnoestructura que modifican el comportamiento, el capital en la nube los ha sacado de la economía de servicios basada en el ser humano y los ha incorporado plenamente a su red de máquinas. Los trabajos que en el marco de la tecnoestructura realizaban los jefes de planta, los publicistas, los expertos en marketing, etcétera, se asignan ahora a algoritmos basados en IA que están totalmente incorporados al capital en la nube“.

Esta contextualización de Varoufakis se concatena con la mirada estratégica que se presentó previamente sobre la cadena de dominación: capitalismo cognitivo+régimen de la información+CGAP+tecno-liberalismo. Este es un armazón muy bien estructurado por el „Bloque Histórico dominante“, explotando el patrimonio intangible por medio de sistemas algorítmicos. Partiendo de Gramsci, esto expresa un proceso hegemónico, donde las plataformas normalizan la explotación bajo la „Construcción de Consenso“ que encausa a la masa hacia lo artificial, lo simulado, lo recreado, lo imitado, para consolidar el sometimiento de lo humano. En tal sentido, desde las clases subordinadas es necesario comprender que una „Guerra de Posiciones“ en la Digitalidad es esencialmente cultural, explicaban Horkheimer y Adorno (1998, 175-176):

„La industria cultural, en suma, absolutiza la imitación. Reducida a mero estilo, traiciona el secreto de éste: la obediencia a la jerarquía social. La barbarie estética cumple hoy la amenaza que pesa sobre las creaciones espirituales desde que comenzaron a ser reunidas y neutralizadas como cultura. Hablar de cultura ha estado siempre contra la cultura. El denominador común „cultura“ contiene ya virtualmente la captación, la catalogación y clasificación que entregan a la cultura en manos de la administración. Sólo la subsunción industrializada, radical y consecuente, es del todo adecuada a este concepto de cultura. Al subordinar todas las ramas de la producción espiritual de la misma forma al único objetivo de cerrar los sentidos de los hombres, desde la salida de la fábrica por la tarde hasta la llegada, a la mañana siguiente, al reloj de control, con los sellos del proceso de trabajo que ellos mismos deben alimentar a lo largo de todo el día, esa subsunción realiza sarcásticamente el concepto de cultura unitaria, que los filósofos de la personalidad opusieron a la masificación“.

Por tanto, el „Bloque Histórico dominante“ para modificar el comportamiento se apoya en una industria tecno/cultural que bombardea comunicativamente al humano individualizado con paquetes algorítmicos de posverdad que aletarga sus capacidades cognitivas.

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

En el estado actual de inmaterialidad social esa „subsunción“ no puede ser tomada como „cultura unitaria“ sino una masificación cosificadora multiplicadora de alienación. Esta explicación de los teóricos críticos es muy interesante porque se articula con la percepción gramsciana, tomando en cuenta que hegemonía cultural/industria cultural son conceptos afines que explican la dominación social, que en la contemporaneidad forjan el control cognitivo con una „imitación“ que trascendió en „simulación digital“ de lo real.

El proceso de cosificación tecno/cultural ha venido perfeccionándose por décadas, el *fordismo cognitivo* abrió el camino pero en la nueva etapa del capitalismo la imitación no era suficiente. Se hizo esencial para la clase dominante el control total del flujo comunicativo para sojuzgar a las masas culturalmente, porque si se logra la hegemonía en ese ámbito las voces disidentes son silenciadas de forma más eficiente. Es decir, nuestra forma de comunicarnos nos puede hacer más fuertes o vulnerables ante la transculturación, así ha sido a lo largo de la historia (quienes dominan lo saben), McLuhan (1996, 100-101) analiza esta transformación en el tiempo:

„El mito griego del alfabeto relata cómo Cadmus, rey que según dicen introdujo el alfabeto fonético en Grecia, sembró dientes de dragón de los cuales brotaron hombres armados. Como cualquier otro mito, éste condensa en un destello intuitivo un largo proceso. El alfabeto supuso poder, autoridad y control a distancia de las estructuras militares. Unido al papiro, el alfabeto fonético dictaminó el fin de la estática burocracia de los templos y del monopolio sacerdotal del saber y del poder. A diferencia de la escritura prealfabética que, con sus innumerables signos, era difícil de dominar, el alfabeto podía aprenderse en unas pocas horas. La adquisición de un conocimiento tan extensivo y complejo como la escritura prealfabética, y su utilización en materiales tan poco maleables como la piedra o el ladrillo, supuso para la casta de los escribas el monopolio del poder sacerdotal. El alfabeto, más sencillo, y el papiro, más barato, ligero y fácil de transportar, efectuaron una transferencia de poder de la clase sacerdotal a la militar. Todo ello está implícito en el mito de Cadmus y de los dientes de dragón, incluida la caída de las ciudades-Estado y la aparición de imperios y burocracias militares“.

La antigua oralidad griega de los bardos rítmicos como sistemas de comprensión auditiva claudicó ante esta poderosa „mordida“ que significó el alfabeto fonético. Es llamativo que lo sucedido en el mundo heleno se replicó en otros entornos donde lo auditivo fue dejado de lado cuando las sociedades alfabéticas se centraron en lo visual, cambiando el patrón comunicativo y de poder. Ciertamente, McLuhan desde su valoración de la *tecnología eléctrica* anticipó la *alfabética algorítmica* que se tejería multisensorialmente, donde datos y emociones terminan entrelazados.

Aquí surge otra complejidad, mientras los bardos usaban el ritmo para grabar relatos en la psique griega los algoritmos artificiales inoculan escenarios

irreales (en una mente acostumbrada a procesar fragmentaria y linealmente). Por lo que la IA puede potenciar los escenarios de posverdad al *crackear* nuestra percepción del mundo.

Tabla 1. Variaciones hegemónicas en los tipos de Comunicación.

Tipo de Comunicación	Base Sensorial	Características Claves	Sensibilidad de la Verdad	Bloque Histórico dominante
Prealfabética (oralidad)	Oído	<ul style="list-style-type: none"> - Colectiva, emotiva, ritualizada. - Memoria auditiva y rítmica. - Mitos como „verdades“ compartidas. 	Alta: La verdad se construye socialmente. El ritmo y la emoción facilitan la manipulación.	Jerarquías Religiosas: Mitificación de lo desconocido.
Alfabética (escritura)	Ojo	<ul style="list-style-type: none"> - Individual, lineal, fragmentada. - Cultura impresa. - Separación contextual. 	Moderada: La escritura permite crítica, pero también distorsión. Depende del acceso a la educación.	Jerarquías Militares/ Intelectuales: Masificación impresa de la mentira.
Posalfabética (digital/ multisensorial)	Multisentido (oído, vista, tacto)	<ul style="list-style-type: none"> - Hiperconexión algorítmica. - Emocionalidad irreflexiva. - Cultura de lo artificial/simulado. 	Muy Alta: Los algoritmos aprovechan la multisensorialidad para amplificar sesgos.	Jerarquías Cognitivas: Imposición de la Posverdad algorítmica.
Fuente: elaboración propia.				

De la tabla previa se pueden sacar varias apreciaciones del desarrollo comunicativo occidental a *grosso modo*, la etapa prealfabética al enfocarse en la tradición oral mantenía un fuerte contenido emocional/emotivo que lo hacía altamente susceptible a una verdad ritualizada, aspecto que era aprovechado por las *Jerarquías Religiosas*. Por su parte, la comunicación alfabética, que se centra en el ojo y la cultura escrita, introduce linealidad, pero es vulnerable a la manipulación textual de las *Jerarquías Intelectuales*. En cambio, en la Digitalidad (posalfabética), la naturaleza multisensorial combina inmediatez, fragmentación y estímulos emocionales, lo que maximiza su exposición a la posverdad estimulada por las *Jerarquías Cognitivas*. Precisamente, los CGAP en la última década del siglo XXI han aprovechado el multisentido e hipersensibilidad de las tecnologías disruptivas para afianzar su dominación, señalaba Gramsci (1999, 387):

„El criterio metodológico en que hay que basar el propio examen es el siguiente: que la supremacía de un grupo social se manifiesta de dos modos, como „dominio“ y como „dirección intelectual y moral“. Un grupo social es dominante de los grupos adversarios que tiende a „liquidar“ o a

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

someter incluso con la fuerza armada y es dirigente de los grupos afines y aliados. Un grupo social puede e incluso debe ser dirigente aun antes de conquistar el poder gubernamental (ésta es una de las condiciones principales para la misma conquista del poder); después, cuando ejerce el poder y aunque lo tenga fuertemente en el puño, se vuelve dominante pero debe seguir siendo también „dirigente““.

Entendiendo que la hegemonía social se configura en dos dimensiones: dominio y dirección intelectual, es visible que la clase dominante se apoya actualmente en la IA para amplificar estas dinámicas: algoritmos y redes sociales refuerzan la dirección intelectual (encauzando „consensos“) y de dominio (censura o desinformación selectiva). Es así, que los CGAP propician narrativas falaces para reforzar el poder hegemónico, disimulando su accionar en una falsa „neutralidad tecnológica“.

5. Conclusiones

El „Bloque Histórico dominante“ en el Capitalismo Cognitivo, se reconfigura como una red de dominación que integra: explotación cognitiva (extracción de datos/subjetividades), exclusión digital (marginalización tecnológica) y control artificial (sujeción de deseos/conductas). Los Centros de Gestión Artificial de la Posverdad (CGAP) (en las que están involucradas plataformas digitales, gobiernos y corporaciones) no solo distorsionan la realidad, sino que fabrican una simulación social donde lo humano es reducido a perfiles cosificados.

La clase hegemónica actual insiste en imponer un „consenso transhumano“ que busca: la superación de la muerte, el envejecimiento selectivo y el desplazamiento de lo biológico. Pero lo anterior, tiene una finalidad: la dominación artificial, la normalización de la cosificación humana y la humanización artificial, todo ello para facilitar la sustitución del colectivo crítico por individualismos algorítmicos inmersos en dinámicas posalfabéticas.

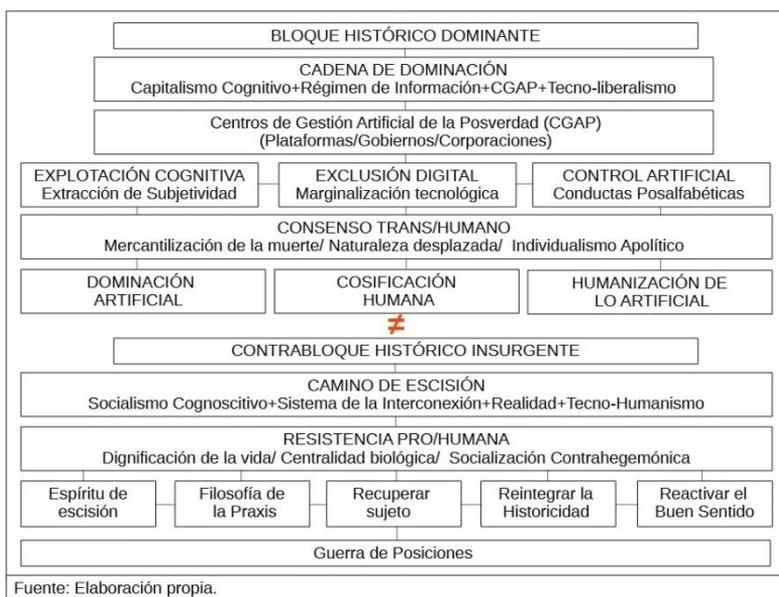
La lucha antes se centraba en los medios de producción, ahora se diversificó, porque se presenta una estructura/superestructura que emerge como cadena de dominación: capitalismo cognitivo+régimen de la información+CGAP+tecno-liberalismo. Ante este contexto, es pertinente que la clase excluida active el germen del espíritu de escisión para romper el „Cautiverio Intelectual“ del „Bloque Histórico dominante“ y propiciar un „Contrabloque Histórico Insurgente“, que planteé una acción política renovada que evite que lo humano quede subsumido en lo artificial.

Todo lo anterior, tiene que ver con transformar la matriz de pensamiento y empezar a dinamizar contra-propuestas sostenidas en una „Filosofía de la Praxis“ que recupere la historicidad del sujeto y se enfrente a temas complejos como la explotación cognitiva, la exclusión digital, la finitud de la muerte, el envejecimiento selectivo, la naturaleza desplazada, la dominación

transhumana, la simulación social, el control artificial, la cosificación humana y los cambios hegemónicos.

Finalmente, es necesario señalar que este estudio ha realizado dos aportes teóricos importantes para analizar el contexto de dominación contemporánea, los conceptos de „Centros de Gestión Artificial de la Posverdad (CGAP)“ y „Contrabloque Histórico Insurgente (CHI)“. Ambas propuestas son parte del desarrollo investigativo del autor y se seguirán profundizando en futuros escritos. En el caso del CGAP es la *pedra angular* de la cadena de dominación, mientras que el CHI es la alternativa de las clases explotadas para recuperar su subjetividad e iniciar el reposicionamiento contrahegemónico.

Esquema 1. Síntesis de la reflexión crítica sobre el uso hegemónico de la IA para la posverdad.



Referencias

Baudrillard, Jean (1991). La guerra del Golfo no ha tenido lugar. Editorial Anagrama.

Benjamin, Walter (2005). Tesis sobre la historia y otros fragmentos. Contrahistorias. La otra mirada de Clío.

Bostrom, Nick (2003). Are You Living in a Computer Simulation? Philosophical

Foth, Marcus y Hearn, Greg. (2007). Networked Individualism of Urban Residents: Discovering the Communicative Ecology in Inner-City

Reflexión crítica sobre el uso hegemónico de la IA para la posverdad

- Apartment Complexes. Information, Communication y Society 10(5). Disponible en: eprints.qut.edu.au/6100/1/6100.pdf
- Gal, Uri** (2025). The race to build a digital god: How our fear of death drives the AI revolution. Disponible: <https://www.abc.net.au/religion/digital-god-how-our-fear-of-death-drives-the-ai-revolution/105505854>
- Gramsci, Antonio** (1981). Cuadernos de la cárcel. Edición crítica del Instituto Gramsci a cargo de Valentino Gerratana. Tomo 1. Cuadernos 1 (XVI) 1929-1930/ 2 (XXIV) 1929-1933. Ediciones Era.
- Gramsci, Antonio** (1981). Cuadernos de la cárcel. Edición crítica del Instituto Gramsci a cargo de Valentino Gerratana. Tomo 2. Cuadernos 3 (XX) 1930/ 4 (XIII) 1930-1932/ 5 (IX) 1930-1932. Ediciones Era.
- Gramsci, Antonio** (1984). Cuadernos de la cárcel. Edición crítica del Instituto Gramsci a cargo de Valentino Gerratana. Tomo 3. Cuadernos 6 (VIII) 1930-1932/ 7 (VII) 1930-1931/ 8 (XXVIII) 1931-1932. Ediciones Era.
- Gramsci, Antonio** (1986). Cuadernos de la cárcel. Edición crítica del Instituto Gramsci a cargo de Valentino Gerratana. Tomo 4. Cuadernos 9 (XIV) 1932/ 10 (XXXIII) 1932-1935/ 11 (XVIII) 1932-1933/ 12 (XXIX) 1932. Ediciones Era.
- Gramsci, Antonio** (1999). Cuadernos de la cárcel. Edición crítica del Instituto Gramsci a cargo de Valentino Gerratana. Tomo 5. Cuadernos 13 (XXX) 1932-1934/ 14 (I) 1932-1935/ 15 (II) 1933/ 16 (XXII) 1933-1934/ 17 (IV) 1933-1935/ 18 (XXXII-IV bis) 1934/ 19 (X) 1934-1935. Ediciones Era.
- Gramsci, Antonio** (1999). Cuadernos de la cárcel. Edición crítica del Instituto Gramsci a cargo de Valentino Gerratana. Tomo 6. Cuadernos 20 (XXV) 1934-1935/ 21 (XVII) 1934-1935/ 22 (V) 1934/ 23 (VI) 1934/ 24 (XXVII) 1934/ 25 (XXIII) 1934/ 26 (XII) 1935/ 27 (XI) 1935/ 28 (III) 1935/ 29 (XXI) 1935. Ediciones Era.
- Han, Byung-Chul** (2021). Infocracia: La digitalización de la Democracia. Barcelona: Taurus.
- Herbert, Edward** (2023). Artificial Intelligence, body image and toxic expectations. Disponible en: <https://www.childrenssociety.org.uk/what-we-do/blogs/artificial-intelligence-body-image-and-toxic-expectations>
- Horkheimer, Max y Adorno, Theodor** (1998). Dialéctica de la Ilustración. Fragmentos filosóficos. Introducción y traducción de Juan José Sánchez. Editorial Trotta.
- Human Rights Watch (HRW)** (2023). Meta's Broken Promises. Disponible en: <https://www.hrw.org/report/2023/12/21/metabrokenpromises/systemic-censorship-palestine-content-instagram-and>
- Joseph, Jeena** (2025). The algorithmic self: how AI is reshaping human identity, introspection, and agency. Sec. Theoretical and Philosophical Psychology Volume 16 - 2025. Disponible en: <https://doi.org/10.3389/fpsyg.2025.1645795>
- Landowski, Eric** (2005). Tres regímenes de sentido y de interacción. Tópicos del Seminario, 14. Julio-diciembre 2005, pp. 137-179. Disponible en: www.redalyc.org/pdf/594/59401408.pdf

- Lawton, George** (2025). What is GenAI? Generative AI explained. Disponible en: <https://www.techtarget.com/searchenterpriseai/definition/generative-AI>
- Maldini, Jorge** (2025). Antonio Gramsci y la Guerra de Posiciones: La Revolución Cultural que No Dispara. Disponible en: medium.com/@jlarabrones1988/antonio-gramsci-y-la-guerra-de-posiciones-la-revoluci3n-cultural-que-no-dispara-1ebb081e448a
- Monga, Ajay** (2023). How Excessive AI Dependence Affects Our Ability to Think and Create. Disponible en: <https://ajaymonga.medium.com/how-excessive-ai-dependence-affects-our-ability-to-think-and-create-24f764dc485b>
- Mcintyre, Lee** (2018). Posverdad. Colección Teorema. Cátedra.
- McKernan, Bethan y Davies, Harry** (2024). 'The machine did it coldly': Israel used AI to identify 37,000 Hamas targets. Disponible en: <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes>
- McLuhan, Marshall** (1996). Comprender los medios de comunicaci3n. Las extensiones del ser humano: Paid3s.
- Nosetto, Luciano** (2017). El sentido com3n en la teor3a del Estado en Antonio Gramsci. Reflexiones con vistas al porvenir sudamericano. Las Torres de Lucca. N3 11 Julio-Diciembre: 131-153.
- Perrella, Alessandro y Maffettone, Ada** (2025). Mortality, belonging, and the paradox of immortality: reflections on the role of AI. AI y SOCIETY, 1-2. Disponible: link.springer.com/article/10.1007/s00146-025-02254-5
- PYMNTS.** (2025). Forever Online: 'Generative Ghosts' Live in the AI Afterlife. Disponible en: www.pymnts.com/artificial-intelligence-2/2025/forever-online-generative-ghosts-live-in-the-ai-afterlife
- Rodr3guez, Olga** (2024). Qu3 hay detr3s de las restricciones y supresiones de contenido sobre Palestina en Instagram y Facebook. Disponible en: https://www.eldiario.es/internacional/hay-detras-restricciones-supresiones-contenido-palestina-instagram-facebook_1_11733003.html
- Rao, Devika** (2024). AI is cannibalizing itself. And creating more AI. Disponible en: <https://theweek.com/tech/ai-cannibalization-model-collapse>
- Satariano, Adam y Mozur, Paul** (2025). The Global A.I. Divide. Disponible en: <https://www.nytimes.com/interactive/2025/06/23/technology/ai-computing-global-divide.html>
- Sibilia, Paula** (2006). El hombre postorg3nico. FCE-Fondo de Cultura Econ3mica.
- Varoufakis, Yanis** (2024). Tecnofeudalismo: El sigiloso sucesor del capitalismo (pp. 126-154). Bilbao: Deusto.
- Ystehede, J3rgen** (2024). AI, reality as simulation and where it leaves us. Disponible en: <https://www.jus.uio.no/ikrs/english/research/news/2024/machine-learning-ai-reality-as-simulation-kaufmann.html>

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA: perspectivas desde la robótica.

Demystifying the attribution of criminal liability for AI actions: perspectives from robotics.

Nathalie Miret González
nattymiret@gmail.com

Resumen

Este artículo tiene por objeto examinar la atribución de responsabilidad penal por las acciones llevadas a cabo por sistemas de inteligencia artificial (IA), con particular énfasis en el ámbito de la robótica, ante el desafío que representa su creciente autonomía. A partir de una metodología cualitativa y un enfoque interdisciplinario, se analizan los fundamentos conceptuales de la IA, sus distintas tipologías y niveles de intervención humana (modelos MITL, MOTL y HOOTL), así como su integración en sistemas robóticos. Se identifican los riesgos asociados a su uso, especialmente cuando participan —de forma autónoma o asistida— en conductas lesivas que podrían encuadrarse en el ámbito penal. Se abordan distintas propuestas doctrinales sobre la responsabilidad penal de la IA, desde su uso como instrumento humano hasta la idea de una posible „personalidad electrónica“. Entre los resultados se destacan las limitaciones del marco legal actual para enfrentar la autonomía creciente de ciertos sistemas, la ambigüedad sobre su imputabilidad, y la necesidad de adaptar las categorías tradicionales del derecho penal. El artículo concluye que, si bien la IA aún no posee conciencia ni estatus jurídico para ser penalmente responsable, su empleo puede implicar consecuencias lesivas que demandan repensar la atribución de responsabilidad penal, especialmente en escenarios de alta autonomía. Se propone avanzar en la creación de marcos regulatorios que incorporen estas nuevas realidades tecnológicas sin eximir a los actores humanos involucrados.

Palabras claves: inteligencia artificial, robótica, responsabilidad penal, personalidad electrónica.

Abstract

This article aims to examine the attribution of criminal liability for actions carried out by artificial intelligence (AI) systems, with particular emphasis on the field of robotics, given the challenge posed by their increasing autonomy. Based on a qualitative methodology and an interdisciplinary approach, we analyse the conceptual foundations of AI, its different typologies and levels of human intervention (MITL, MOTL and HOOTL models), as well as its integration in robotic systems. It identifies the risks associated with their use, especially when they participate - autonomously or assisted - in harmful conduct that could fall within the criminal sphere. Different doctrinal proposals on the criminal liability of AI are addressed, from its use as a human instrument to the idea of a possible „electronic personality“. The results highlight the limitations of the current legal framework to deal with the growing autonomy of certain systems, the ambiguity about their imputability, and the need to adapt the traditional categories of criminal law. The article concludes that, although AI does not yet have the awareness or legal status to be criminally responsible, its use can entail harmful consequences that require rethinking the attribution of criminal responsibility, especially in scenarios of high autonomy. It is proposed to advance in the creation of regulatory frameworks that incorporate these new technological realities without exempting the human actors involved.

Keywords: artificial intelligence, robotics, criminal liability, electronic personality.

1. Entre el ideal de Robotina y el síndrome de Blade runner: notas introductorias

El 23 de septiembre de 1962 se estrenó la serie animada Los Supersónicos (*The Jetson*), de William Hanna y Joseph Barbera, presentando a la familia Jetson, en su mundo futurista colmado de tecnología de avanzada, autos voladores, vacaciones espaciales y robots domésticos, impactando de manera significativa en la sociedad y la forma de concebir la robótica. La serie mostró una convivencia cotidiana con robots, normalizando la idea de que las máquinas podrían integrarse en la vida diaria y mantener una interacción sana con los seres humanos. Quizás la figura más icónica en este sentido lo fue Robotina, quien se convirtió en el modelo de los asistentes robóticos anticipando la idea de sus aplicaciones en el servicio en el hogar.

Los Supersónicos no solo fueron entretenimiento, sino que también ayudaron a moldear la visión del futuro y a inspirar desarrollos tecnológicos que hoy

forman parte de la vida cotidiana. No representan una distopía sombría ni una utopía idílica, sino un mundo donde el capitalismo y el espíritu empresarial todavía existen y la tecnología no ha cambiado elementos fundamentales de la naturaleza humana (Tucker, 2011).

En 1968, Philip K. Dick publicó su novela *¿Sueñan los androides con ovejas eléctricas?*, en la que se basó, cerca de veinte años más tarde, la película *Blade Runner*, dirigida por Ridley Scott. En la misma, se muestra también un mundo futurista y distópico, pero con un matiz mucho más oscuro que *The Jetsons*. La inteligencia artificial aplicada a la robótica ya es un hecho, y han sido desarrollados un sin número de modelos de robots androides llamados replicantes, a los que se le han atribuido emociones humanas y conciencia. La diferencia es que, mientras los robots de *Los Supersónicos* son acogidos con cariño, los de esta entrega de Ciencia Ficción son sometidos a trabajos peligrosos en colonias espaciales y cazados para su exterminio una vez que se rebelan en busca de una mejor vida.

Esta obra presenta una cara distinta de los avances tecnológicos. Mientras la serie de dibujos animados se centra en las ventajas de la robótica y la inteligencia artificial, la película abre la puerta a dilemas morales importantes al cuestionar si la conciencia y las emociones son exclusivas de los seres humanos, si los robots deberían tener derechos o responsabilidades. Como *Los Supersónicos*, fue capaz de vaticinar el futuro, y lo que hoy se conoce como Síndrome de *Blade Runner* (Navajas, 2020), en cuanto a la despersonalización del hombre o la humanización de las máquinas.

„Descartes creía que los organismos vivos son máquinas biológicas, sometidas como cualquier otra entidad física a las leyes del universo. Los seres humanos, en cambio, seríamos diferentes, poseedores de libre albedrío. Pero Rick Deckard, (...), tenía serias dificultades para distinguir a un «replicante», un robot biológico, de un ser humano, y así no equivocarse a la hora de «retirar» (eufemismo de matar) a las «entidades electrónicas». Cuando Isaac Asimov se planteó cómo sería la relación entre humanos y máquinas, comprendió que a medida que los robots se hiciesen más complejos y les encargásemos más tareas antes en manos humanas, habría que dotarles de ciertas reglas morales «innatas» o programadas de serie. De lo contrario, podrían ser peligrosos en su relación con los seres humanos“ (Navajas, 2020)

Constituye una tendencia en los estudios recientes iniciar los análisis relacionados con la inteligencia artificial (IA) y su impacto en las relaciones sociales y el desarrollo, haciendo referencia a la llamada 4ta revolución industrial (Serrano, 2024) en el contexto de la sociedad del riesgo (Beck, 1997); y es que no resulta ociosa la remisión a ambas perspectivas en esta sede, considerando que, en efecto, los tiempos actuales han sido bautizados como la era de la inteligencia artificial, y una de las mayores preocupaciones gira en torno a las ventajas o desventajas que pueda ofrecer a la especie

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

humana, así como el delimitar un marco legal que permita regular su uso y explotación en condiciones éticas, seguras y sostenibles.

La inteligencia artificial ha surgido rápidamente como una herramienta transformadora, reconfigurando industrias y sociedades en todo el mundo al ofrecer soluciones innovadoras a problemas complejos a través de la simulación de la cognición humana. Uno de los campos más fascinantes en los que se aplica, es precisamente la robótica, integrándose en prácticamente todos los sectores de interacciones sociales.

Sin embargo, de conjunto con sus innumerables beneficios, ha planteado importantes cuestiones legales, éticas y sociales. Elementos como electrodomésticos inteligentes y sistemas de control por voz que fueron imaginados en Los Supersónicos, hoy son una realidad; pero también lo es la preocupación creciente acerca de los desafíos que presentan las máquinas en cuanto a su desarrollo exponencial y el acercamiento a la total autonomía y el incremento de los riesgos en lo que respecta a su participación en actividades delictivas.

Uno de los problemas más urgentes es la falta de un marco legal claramente definido que aborde los desafíos únicos que plantea la robótica inteligente en esta sede. A diferencia de las herramientas o tecnologías tradicionales, estos sistemas de IA pueden operar de manera autónoma, tomando decisiones y llevando a cabo acciones sin supervisión humana directa en algunos casos. Esta autonomía introduce un dilema legal complejo: cuando un sistema de IA está involucrado en un delito, ¿quién debe ser considerado responsable? ¿Es la IA misma, sus creadores o sus usuarios?

El potencial de que los robots inteligentes se utilicen de maneras que perjudiquen a individuos o a la sociedad en general resalta la urgente necesidad de que los cuerpos legislativos y la jurisprudencia intervengan y llenen este vacío legal. A medida que los sistemas de IA se vuelven más sofisticados y autónomos, hay un riesgo creciente de que puedan ser explotados para fines criminales o que sus acciones puedan, sin querer, conducir a resultados perjudiciales. Esta realidad subraya la necesidad de legislaciones penales modernas que reflejen los avances en la tecnología de la IA. Tales normativas deben estar diseñadas para gestionar las complejidades propias de este escenario, asegurando que la responsabilidad esté claramente establecida y que se pueda hacer justicia cuando la IA esté involucrada en actividades criminales (Hammouri y otros, 2024).

2. Inteligencia artificial y robótica. Del laboratorio al mundo real

2.1. Inteligencia artificial. (De)construyendo significados

Existe cierto consenso en la doctrina (Valls, 2022; De la Cuesta, 2019; Pino y otros, 2001) acerca de la dificultad y variabilidad subyacente en cuanto a la conceptualización de inteligencia artificial, en tanto cualquier definición que se pretenda elaborar, responde a los presupuestos y objetivos de la disciplina desde la que se analice. En este orden, Amador (1996, 15) considera que ello obedece a dos razones fundamentales: por un lado, la creencia de que la inteligencia resulta exclusiva de la especie humana y la correspondiente tendencia a rechazar la probabilidad de que una máquina pueda desarrollar características o capacidades mentales similares; y por el otro los sesgos propios del concepto mismo de inteligencia.

Sin embargo, proveniente del latín *intelligere*, el término hace referencia a la capacidad de comprender las relaciones existentes entre los hechos y las cosas, así como para resolver problemas a partir de la evaluación de distintas opciones y la adaptabilidad a nuevas situaciones (García, 2020, 157-161); por lo que, de hecho, puede hablarse de inteligencia no solo en la especie humana sino también en la especie animal e incluso, en cierto grado, en las máquinas. Inteligencia, en definitiva, no significa lo mismo que sabiduría, conciencia, o memoria.

La sabiduría constituye una cualidad que supera los límites de la inteligencia, permitiendo la aplicación reflexiva y ética del conocimiento, teniendo en cuenta las experiencias vividas, las emociones y el impacto de cada decisión. Así mismo la conciencia implica el percatarse de los pensamientos, la existencia del yo y la acción, en un sentido, y la capacidad de distinción entre el bien y el mal y el reconocimiento de deseos y necesidades en el otro (Damasio, 2019). La memoria por su parte, se refiere al almacenamiento y retención de datos, y la recuperación de información; permitiendo recordar experiencias, aprender de ellas y reutilizarlas en el futuro. Por ello, si bien es dable afirmar que una máquina no posee conciencia o sabiduría, forzosamente habría que admitir que sí pueden tener memoria, e incluso inteligencia, independientemente a si se considera como un atributo natural o artificialmente construido.

La particularidad distintiva en este caso proviene del adjetivo mismo que la describe, el cual le otorga por sí solo una connotación específica, al describirla como no-natural y creada por el ingenio humano para simular su propio intelecto. Morales afirma que la inteligencia artificial tiene como objeto de estudio los fundamentos en que se basa el desempeño inteligente del ser humano en función de asimilar cómo estos utilizan el lenguaje, los procesos de aprendizaje e inferencias, comprensión de fenómenos, las formas de percepción, así como los mecanismos de adquisición, aplicación y transferencias de conocimientos para la planeación, ejecución y evaluación de planes de actuación; con la finalidad de reproducirlos y aplicarlos en el funcionamiento y desarrollo de las máquinas (Morales, 1997).

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

La inteligencia artificial se traduce en la ciencia y la ingeniería de hacer máquinas y programas informáticos inteligentes (McCarthy, 2007), siendo este el atributo que le otorga al sistema la capacidad de interpretar de manera correcta los datos externos, aprender de ellos y emplear el conocimiento adquirido para resolver problemas y realizar tareas concretas a través de una adaptación flexible (Kaplan y Haenlein, 2018); con lo cual la máquina puede llegar a simular el pensamiento y comportamiento humano.

Desde esta óptica, el concepto de inteligencia artificial se inserta en el espectro de las ciencias computacionales y la ingeniería, vertebrando, de acuerdo con el criterio de Pelet (2024), dos perspectivas interconectadas. Por una parte, se enfoca en el análisis de los procesos cognitivos para la comprensión teórica y práctica del intelecto que pueda llegar a demostrar una máquina y las distintas actividades que realiza; mientras que, por el otro, tiene como objetivo precisamente la creación de máquinas y sistemas que sean capaces de poseer inteligencia y llevar a cabo tareas en las que usualmente los humanos tendrían un mejor desempeño, siendo capaces en todos los casos, de modificar el mundo exterior. Se identifica, en síntesis, con los razonamientos algorítmicos y las representaciones simbólicas del conocimiento, a través de la estructura que permite a un ordenador ser inteligente.

Como elementos comunes a todas las definiciones ofrecidas, se sistematizan los siguientes (Samoili, y otros, 2020):

- Percepción del entorno, incluyendo la consideración de la complejidad del mundo real.
- Procesamiento de información, recopilación y análisis de datos
- Toma de decisiones a partir del aprendizaje y el razonamiento
- Realización de acciones y ejecución de tareas con cierto grado de autonomía
- Consecución de objetivos específicos.

La problemática con estas definiciones radica en que, desde el punto de vista tecnológico, pueden resultar suficientes para la comprensión del fenómeno, empero, desde el punto de vista social y jurídico, no logran delimitar una serie de cuestiones claves en cuanto al alcance, funcionamiento, aplicabilidad e impacto de la inteligencia artificial en la sociedad. Por tanto, el análisis de su trascendencia y la resolución de los disímiles conflictos que podrían generarse a partir de su instrumentalización, se sitúa en una zona gris que demanda una mirada más profunda a sus fundamentos conceptuales.

En este orden, Valls (2022) apunta que el hecho de que una máquina sea inteligente, implica que es a su vez capaz de pensar, razonar y actuar con inteligencia de manera idéntica a un ser humano en determinadas

circunstancias. No obstante, se considera que esta mirada merece una reinterpretación crítica; por cuanto si se hace referencia al proceso de pensamiento como habilidad cognitiva de reflexión consciente, que involucra emociones y experiencias subjetivas, evidentemente habría que admitir, sustentado en los aspectos ya analizados sobre la conciencia y la sabiduría, que las máquinas no son capaces de pensar en el sentido en que lo haría un humano.

En este contexto, esa habilidad de pensar se traduce en la capacidad de procesamiento de información, resolución de problemas y generación de respuestas basadas en el análisis de los datos que posee, algoritmos y modelos matemáticos, además de la adaptabilidad a diferentes situaciones; todo lo que dibuja una suerte de razonamiento lógico en la caracterización de su comportamiento, ideado para imitar el de un ser humano, sin incorporar la carga de subjetividad propia de aquel, más allá de aquella que quedará incorporada en los mecanismos de programación efectuados por su desarrollador.

Los sistemas de inteligencia artificial no dejan de ser elementos de software o hardware diseñados por seres humanos, si bien, una vez que se les presenta un objetivo complejo, pueden actuar en la dimensión física o digital, percibir su entorno mediante la adquisición e interpretación de datos estructurados o no estructurados, razonar sobre el conocimiento y el procesamiento de la información derivada de esos datos, e identificar y aportar las mejores soluciones para darle cumplimiento (Morillas, 2023; Comisión Europea, 2019).

Como sistemas de software inteligentes con especiales implicaciones a nivel social, la Comisión Europea (2019) ha identificado a los asistentes de voz, software de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz, softwares generativos de contenido como textos u otros materiales audiovisuales y señala como hardware inteligentes a los robots avanzados, drones o coches autónomos, entre otros. En todos los casos, se enfatiza en que su actuación se pone en marcha a partir de la solicitud realizada por un ser humano, y que sus predicciones, recomendaciones o decisiones pueden influir de manera directa en los entornos con los que el sistema interactúa.

Coinciden Pelet (2024) y Momblanc (2024) en que esta definición resulta más acertada en el contexto socio-jurídico, pues delimita la naturaleza de la inteligencia artificial como un producto de la creación humana, un instrumento que, en principio, no debe realizar ninguna tarea que no sea predeterminada por su creador, desarrollador o usuario. Su actuación se basa en los datos que previamente le fueron suministrados, y detrás de los cuales, nuevamente, se halla la intervención directa de un ser humano, por lo que no gozan de total autonomía e independencia. Ello conlleva a que en su ámbito de aplicación práctica resulte posible definir las medidas de cuidado que deberán tomarse

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

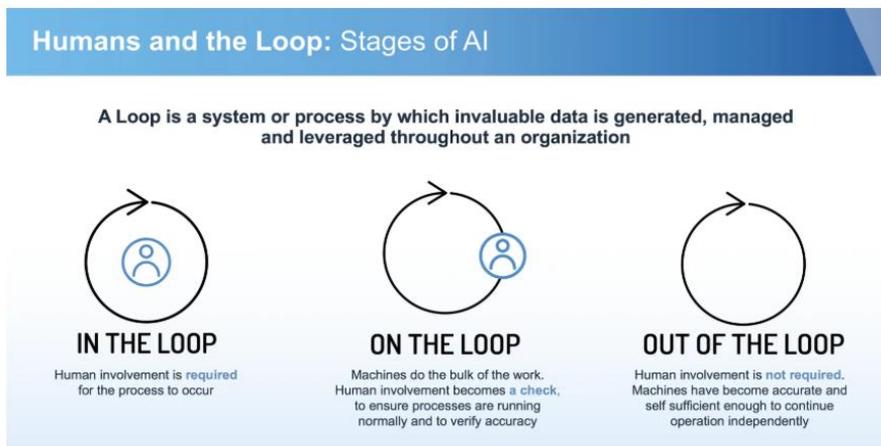
para evitar el impacto negativo que su explotación pueda representar para la sociedad.

El alcance de los sistemas de inteligencia artificial en este orden, cubre todo un rango de posibilidades que van desde la amplia gama de instrucciones que sean capaces de ejecutar, hasta el grado de autonomía con que aquellas sean llevadas a cabo, en relación con los niveles de dependencia que presenten respecto al ser humano. En correspondencia con ello, Miró (2018) sostiene que lo verdaderamente importante no es la determinación cuantitativa de sus funcionalidades específicas y las habilidades concretas que puedan poseer, sino el análisis cualitativo de su capacidad real de razonamiento y el nivel de equivalencia establecido entre las complejidades propias del procesamiento de información realizado por la máquina y aquel que desarrolla el cerebro humano, considerando el grado de independencia en su autogestión y aprendizaje; aspecto que resulta medular para la comprensión de los niveles de autonomía, el impacto y las responsabilidades atribuibles en cada caso.

2.1.1. Hacia una tipología de inteligencia artificial

El grado de intervención e interacción entre el ser humano y la máquina o sistema inteligente, permite establecer una diferenciación entre los modelos de inteligencia artificial existentes, evaluados por Miró (2018) como *Man in the loop*, *Man on the loop*, o *Man out of the loop*, los que se representan gráficamente a continuación:

Figura 1.
Representación de modelos de IA de acuerdo con la interacción hombre-máquina.



Fuente: Thakur, Raman (2025)

En primer lugar, el modelo *Man in the loop*, también conocido como MITL, se caracteriza porque la inteligencia artificial requiere de la intervención humana

en lapsos de tiempo regulares para poder llevar a cabo sus tareas, de manera que carece de total autonomía e independencia. Al actuar como un sistema de apoyo, la supervisión humana es fundamental para la validación y ajuste de las respuestas que ofrece. Su principal utilidad y aplicación está encaminada a campos en los que no resulta confiable una completa automatización en pos de la precisión y la preservación de la ética. Aunque en estos casos el espectro de libre actuación de la IA se encuentra limitado, sin dudas es uno de los que ofrece una mayor seguridad en cuanto a los resultados de su explotación, por cuanto no se desliga de la correspondiente actuación humana y el cuidado debido, de manera que también resulta el menos problemático para delimitar responsabilidades.

Por su parte, el modelo *Man on the loop* (MOTL) se refiere a sistemas que poseen la capacidad de actuar por sí mismos gracias a una programación previa, aunque el ser humano puede modificar o interrumpir la actividad que la inteligencia artificial desarrolla en cualquier momento; por lo que posee en este caso una autonomía relativa bajo supervisión ocasional que permite la intervención ante situaciones críticas. La finalidad de este modelo, subyace en el equilibrio entre las ventajas de la automatización y el control humano para evitar errores graves.

Finalmente, el modelo *Man out of the loop* (HOOTL)¹, resulta quizás el más polémico. La inteligencia artificial una vez activada, se desarrolla en estos con completa autonomía, ya que puede actuar por sí misma de manera independiente, durante determinados intervalos de tiempo, sin que el ser humano tenga ninguna influencia sobre sus acciones o decisiones. Ello implica a su vez importantes desafíos éticos en cuanto al impacto de su funcionamiento y la atribución de responsabilidad por las consecuencias que de él se deriven.

Cada uno de estos modelos presenta sus ventajas, desventajas, riesgos e implicaciones éticas, los que se señalan a continuación:

Tabla 1. Análisis comparativo entre los modelos MITL, MOTL, y HOOTL:

Modelos	Ventajas	Desventajas	Riesgos	Cuestiones éticas
Man in the Loop	Mayor control humano sobre las decisiones de la IA. El ser humano participa directamente en cada decisión.	Procesos más lentos debido a la intervención humana constante. Menor eficiencia.	Posibilidad de sesgos humanos en la toma de decisiones.	Responsabilidad clara del humano en cada decisión.

¹ **Nota de la Autora:** Las siglas obedecen a la calificación de Human out of the loop, como también se le conoce.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

	Reducción de errores graves gracias a la supervisión directa. Mayor responsabilidad y trazabilidad.	Dependencia de la disponibilidad y capacidad del operador.	Fatiga del operador, lo que puede afectar la calidad de las decisiones.	Transparencia en el proceso de toma de decisiones.
	Mejora la confianza del usuario.	Puede limitar la escalabilidad de los sistemas automatizados	Menor eficiencia en tareas que podrían automatizarse completamente	Protección contra errores graves o decisiones injustas
Man on the Loop	Equilibrio entre autonomía de la IA y supervisión humana.	Posible falta de intervención humana en momentos críticos.	Dificultad para intervenir en tiempo real si la IA comete errores.	Necesidad de transparencia en el funcionamiento de la IA.
	Mayor eficiencia en comparación con MITL.	Dependencia de la calidad de los algoritmos de IA.	Posible pérdida de control si el sistema falla.	Definición clara de cuándo y cómo los humanos pueden intervenir. Dificultad para asignar responsabilidad.
	Capacidad de intervención en caso de fallo	Supervisión limitada en escenarios de alta velocidad de decisión.	Riesgo de sesgos en los datos utilizados por la IA	Regulaciones para evitar decisiones automatizadas perjudiciales. Potencial deshumanización parcial del proceso decisorio
Man out of the Loop	Máxima eficiencia y velocidad en la toma de decisiones.	Falta de supervisión humana puede llevar a decisiones erróneas.	Posibles consecuencias graves si la IA toma decisiones incorrectas.	Debate sobre la responsabilidad en decisiones tomadas por IA.
	Respuesta rápida en tiempo real y escalabilidad	Dificultad para corregir errores en tiempo real.	Problemas de seguridad en sistemas autónomos sin intervención humana. Opacidad algorítmica.	Necesidad de regulaciones estrictas para evitar abusos y discriminación.
	Reducción de errores humanos en tareas repetitivas.	Riesgo de pérdida de control sobre el sistema.	Dilemas éticos en aplicaciones militares y de seguridad.	Riesgo de pérdida de control sobre sistemas autónomos avanzados.

Fuente: Elaboración propia, 2025.

Se ha valorado por la doctrina una pauta adicional en cuanto a la clasificación tipológica de la IA según su capacidad de imitar las características cognitivas humanas, estableciéndose tres categorías fundamentales: inteligencia artificial débil o estrecha, también denominada *narrow IA*, con un limitado espectro de habilidades específicas; la general o fuerte, *strong or general IA*, con habilidades similares a las que poseen los seres humanos; y la superinteligente, cuyas capacidades superan las humanas.

De acuerdo con Porcelli (2020) la inteligencia artificial débil está únicamente enfocada en la realización de tareas específicas consideradas estrechas. Es capaz de llevar a cabo tareas inteligentes, pero no posee la capacidad de pensar, en el sentido en que ya se ha expresado que podría hacerlo una máquina, de manera casi idéntica a un humano. Por ello, sería fácilmente reconocible que ante determinadas interacciones el usuario no está en presencia de un igual, sino que su interlocutor es precisamente un sistema informático.

Realiza de manera automática una tarea concreta, a partir de clasificaciones basadas en parámetros predeterminados e invariables sobre los que fue programada, los que se van puliendo según la cantidad de datos que posee, pero no exceden esa programación original. Estas tareas son rutinarias y técnicas en función de los gustos del usuario, y pueden ser tan simples como clasificar correos electrónicos en deseados o spam (Porcelli, 2020), o algo más complejas como aquellas dirigidas al reconocimiento facial, asistentes de voz o conducción (Escott, 2017). Este es el tipo de IA que se representa en todos los modelos desarrollados en la actualidad y que han alcanzado mayor éxito, a pesar de que su proceso de aprendizaje no puede desdoblarse y convertir cada experiencia en un nuevo conocimiento aplicable a otra problemática con un enfoque distinto.

Al respecto, Morales (2021) establece un paralelismo entre el desempeño de un ser humano y un ordenador frente a un juego de damas y uno de ajedrez, explicando que si bien existen programas diseñados para jugar a este último en nivel Gran Maestro o Experto, no podrían utilizar los mismos datos y algoritmos para jugar de la misma forma a las damas, ni siquiera para realizar movimientos estratégicamente planeados en el propio juego, ya que sería necesario el diseño y ejecución de una programación completamente distinta, pues está imposibilitado de adaptar sus conocimientos para el primero al segundo. Por el contrario, un ser humano podría aprovechar lo que ya sabe, transformarlo y aprender algo nuevo a partir de la experiencia previa.

La autora señala que la IA débil, por tanto, se refiere a programas que no tienen estados mentales, a pesar de realizar tareas específicas de manera inteligente, buscar soluciones a fórmulas lógicas con múltiples variables, razonar deductivamente o aprender de manera inductiva y tomar decisiones. A diferencia de los humanos, la información no es procesada mediante la

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

activación coordinada de redes neuronales, sino a través de operaciones aritmético-lógicas, de lectura, escritura, registros y control de flujo secuencial (Morales, 2021).

Tratándose incluso de un sistema experto, como los conocidos por Knowledge Based System, capaces de procesar y memorizar grandes volúmenes de información, aprender y razonar en situaciones complejas, comunicarse con los humanos u otros sistemas similares, y tomar decisiones que además pueden explicar y justificar en ámbitos en los que pueden sustituir la intervención humana y actuar de manera fiable, no dejan de ser programas de inteligencia artificial estrecha. Al carecer de conciencia y sentido común, sus capacidades cognitivas quedan limitadas a la programación base (Amador, 1996).

Los sistemas de IA débil suelen operar bajo el modelo MITL, donde el hombre tiene una participación más activa en todo su proceso de funcionamiento, y generalmente se trata de sistemas menos autónomos; y aunque no son los más frecuentes, pueden también actuar, en casos de sistemas semi-autónomos o más avanzados, bajo el modelo MOTL. Por el contrario, no suele ser aplicable a este tipo de IA el modelo HOOTL, ya que al operar de manera totalmente autónoma implica riesgos importantes en cuanto al impacto social de sus decisiones, control y adjudicación de responsabilidad.

En contraposición la inteligencia artificial fuerte o general, en el plano teórico, tiene capacidades cognitivas comparables o superiores a las de un ser humano, en cuanto supone una capacidad de abstracción, reflexión, creatividad e improvisación que la llevarían a integrar de hecho una mente en sí misma. Los criterios en la doctrina son unánimes (Morales, 2021; Pelet, 2024; Porcelli, 2020) en cuanto a considerar que de hecho la existencia de este tipo de IA en la actualidad es aún una utopía, si bien existen sistemas y aplicaciones tan avanzados, que podrían constituir sus primeros gérmenes.

La IA fuerte es capaz de replicar el comportamiento y la inteligencia humana, aprender de manera autónoma, y aplicar esa inteligencia para resolver problemas de la manera en que lo haría una persona, sin obedecer a un patrón previo de programación. De acuerdo al criterio de Del Rosal (2023), lograr una IA de este tipo, implicaría crear una máquina con conciencia a un nivel verdaderamente humano, más allá de una simple réplica.

La concepción de la IA fuerte, se considera como el paso precedente a la creación de una superinteligencia, tercera categoría sobre la que se han esgrimido interesantes criterios. En palabras de Bostrom (2014), poseería un intelecto y desempeño cognitivo muy superior al de los cerebros humanos, incluyendo las habilidades sociales, incorporando un estadio de conciencia que les permitiría reconocerse a sí mismos como seres pensantes. En su

máxima expresión evolutiva, este tipo de IA no se limitaría a reproducir las emociones y experiencias humanas, sino que evocaría las suyas propias, necesidades, deseos y creencias (Escott, 2017), de manera que sería prácticamente imposible distinguirlas de los seres humanos, encontrándose un paso más cerca de hacer real la paradoja de *Blade Runner* y su representación de los replicantes como robots humanizados.

A partir de estos elementos, Hintze distingue cuatro grandes grupos en los que podría clasificarse la inteligencia artificial de acuerdo a su nivel de complejidad (Hintze, 2016):

I. Máquinas reactivas:

En este grupo se sitúan los tipos más básicos de inteligencia artificial, los que no tienen la capacidad de crear recuerdos ni de utilizar experiencias pasadas para justificar decisiones actuales. Como el ejemplo paradigmático, el autor sitúa a *Deep Blue*², y explica que, a pesar de su impresionante capacidad de cálculo, habilidades para predecir los movimientos futuros de su oponente en una partida de ajedrez, y la elección de las mejores jugadas entre muchas posibilidades, no tiene conciencia del mundo más allá del objetivo para el que fue programada. Como *Deep Blue*, este tipo de IA no reconoce la realidad exterior ni puede funcionar fuera de las fronteras de las tareas específicas que le fueron asignadas. De ahí que sus interacciones con el entorno demostrarán siempre el mismo comportamiento ante situaciones similares. Ello, si bien asegura por un lado la confiabilidad del sistema, por el otro limita su rendimiento, al no ser capaces de responder ante las modificaciones en los estímulos externos variables. Al mismo tiempo, son las más fáciles de comprender para los seres humanos ya que hace gala del funcionamiento tradicional de una computadora a la que se le proporcionan una serie de datos y debe procesarlos para obtener un resultado satisfactorio.

II. Memoria limitada:

Su denominación hace referencia a la capacidad que poseen estos sistemas para registrar experiencias pasadas y aplicarlas en las decisiones actuales, aunque no almacenan recuerdos. Sus principales aplicaciones van desde la industria automotriz y los vehículos autónomos hasta la IA integrada en los dispositivos móviles y asistentes de voz, siendo capaces de entender una representación del mundo. Aprenden de manera automática, y en base a la

² **Nota de la Autora:** Deep Blue fue una supercomputadora desarrollada por IBM, famosa por ser el primer sistema de inteligencia artificial en derrotar a un campeón mundial de ajedrez bajo condiciones de torneo estándar. Evaluaba hasta 200 millones de posiciones por segundo, lo que le permitía analizar jugadas con una profundidad impresionante. Marcó un hito en la historia de la IA cuando venció, en 1997, a Garry Kasparov, campeón mundial de ajedrez, en un enfrentamiento de seis partidas.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

información de base, incorporan nuevos datos para mejorar su rendimiento (Porcelli, 2020).

III. Teoría de la mente:

Este grupo representa la bandera que marca un antes y un después en el ámbito de la inteligencia artificial, entre las máquinas que hasta el momento existen y aquellas que serán creadas en el futuro. Este tipo de IA no solo será capaz de representar y entender el mundo, sino de conformar sus propias representaciones de otros agentes o entidades, comprendiendo la existencia de las personas, criaturas y objetos y que estos poseen emociones y pensamientos que afectan su comportamiento. Ello les permitiría tener interacciones sociales más precisas y tomar decisiones conscientes, ajustando su conducta al entorno. Esa capacidad de procesar emociones, supondría el primer paso para integrar en sociedad a la inteligencia artificial, como parte de ella, y no como una simple máquina.

En opinión de Porcelli (2020), actualmente este tipo de IA se encuentra en fase experimental, no obstante, ya existen procesos capaces de detectar estados de ánimo en las personas por los gestos de su cara, los matices de la voz, gritos, insultos y otros indicadores de conducta humana.

IV. Autoconscientes

Sería este el estadio superior en la cadena evolutiva de la inteligencia artificial, con la mayor proximidad a los seres vivos. No se estaría hablando solo de que la máquina podría crear y entender representaciones del mundo, sino que sería capaz de representarse y ser autoconsciente, procesando las emociones de los otros y expresando las suyas.

Desde este esquema, los grupos I y II se identifican con la IA estrecha, y única que a día de hoy ha sido creada y explotada por el hombre. El grupo III hace referencia a la IA general, y en un plano utópico, el grupo IV evoca a una superinteligencia que hoy se tiende a identificar como mito. En este orden, se reconoce que existen sistemas que han logrado un desarrollo tal, que se han llegado a considerar como manifestaciones incipientes de inteligencia artificial fuerte, desde sistemas generativos, hasta las más revolucionarias aplicaciones en la robótica.

Como los ejemplos emblemáticos, se citan a *ChatGPT* y *Gemini*, sistemas de procesamiento de lenguaje natural avanzado, sin embargo, no poseen conciencia, emociones ni razonamiento autónomo; y robots como Sophia, humanoide capaz de entablar conversaciones con personas, y Boston Dynamics, quien destaca por su agilidad y movilidad para ejecutar tareas físicas de elevada complejidad, aunque en el primer caso, solo simula las

expresiones humanas y hace uso de frases programadas, mientras que en el segundo, no puede tomar decisiones por sí mismo. De ahí que, sin llegar a constituir verdaderas muestras de inteligencia artificial fuerte, son únicamente manifestaciones avanzadas de IA débil.

En el mismo sentido, la mayoría de los sistemas de inteligencia artificial que actualmente se hallan en funcionamiento, operan bajo los modelos de *Man in the Loop* o *Man on the Loop*, siendo extremadamente raro encontrar sistemas *Man out of the Loop* verdaderos y complejos, más allá de muestras en entornos controlados como fábricas automatizadas o comercio de alta frecuencia, pues existe cierta tendencia a mantener a los humanos como agentes supervisores en algún grado de la actividad que realizan las máquinas. El reto en desarrollar un modelo HOOTL real radica en que requeriría una robustez en sus redes neuronales artificiales propia de la IA general. Al mismo tiempo, ello constituye su principal riesgo, al demandar unos límites operativos muy estrictos, dado que se trataría de sistemas que actuarían sin control ni supervisión humana, lo que además afectaría la transparencia (Thakur, 2025).

La controversia fundamental en esta sede se halla en las aplicaciones que en el área de la robótica tiene la IA, y en el desarrollo de los modelos y tipologías comentadas en ese campo, dado que los avances en la creación de robots inteligentes para el desarrollo de disímiles tareas en los ámbitos más variados ha sido vertiginoso, llegando a integrarse, aun cuando de manera especulativa, en la industria armamentista y de seguridad, donde ya se habla de drones completamente autónomos con aplicaciones prácticas muy precisas en contextos de guerra.

Aún y cuando la IA general no es una realidad, el número de sistemas inteligentes que se programan son cada vez más complejos y de mayor envergadura (Stanford University, 2023). Esta preocupación ha sido también expresada por el Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías, al considerar que los sistemas más avanzados son menos transparentes, pues han dejado de ser programados en gran parte por seres humanos, lo que puede desembocar en que las acciones llevadas a cabo por robots u otras manifestaciones de IA se conviertan en indescifrables ante la imposibilidad de determinar cómo generan sus resultados más allá de los algoritmos iniciales (Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías: Comisión Europea, 2018).

Es por ello que se sitúa como una de las principales preocupaciones y tema de debates en la actualidad la determinación de los parámetros para la atribución de responsabilidad ante la actuación de un sistema de este tipo, sobre todo cuando se estime ilícita y de ella se derive un peligro o daño hacia bienes jurídicos penalmente protegidos por su significación para la sociedad. Aunque no debe ignorarse que, en palabras de Momblanc (2024), el sistema

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

de inteligencia artificial se halla en nuestros días „embrionariamente ligado al ser humano y no goza de absoluta autonomía“, el asunto se torna en extremo problemático en aquellos casos en que su evolución implique una separación diametral entre la intervención de la persona y la de la máquina, dificultando así la imputación de las consecuencias lesivas al programador, desarrollador o usuario.

2.2. De la IA en la Robótica: complejidades conceptuales

Una de las ramas más fascinantes de la tecnología moderna es, sin dudas, la robótica, tanto desde sus representaciones en la literatura y el cine de ciencia ficción, como en la vida real. Constituye un campo innovador con potencial para la aplicación de la IA en el desarrollo de máquinas capaces de interactuar y transformar el mundo, sobre todo a partir de los avances alcanzados en las últimas décadas en el área de la automatización. Desde las tareas domésticas más simples, hasta la optimización de procesos de producción en fábricas y complejos sistemas de seguridad, la robótica ha logrado infiltrarse en muchos de los escenarios cotidianos de interacciones sociales importantes.

El origen del término robot, se halla en el idioma checo, en la palabra *robota*, cuyo significado es trabajo forzado o mano de obra; aunque en la actualidad no es exactamente esta la connotación que se le otorga, ya que, si bien muchos robots han sido diseñados para llevar a cabo tareas pesadas, hoy apuntan a toda una variada gama de aplicaciones (Rivera, 2023). Aunque sus antecedentes han sido rastreados desde el Siglo I a.C, con sus primeras apariciones en literatura -espacio donde ha sido un tema recurrente-, y la juguetería japonesa, fue Karel Čapek en 1920, en su obra *Rossum's Universal Robots*, quien por primera vez hizo uso del vocablo para referirse a máquinas humanoides que en su comportamiento resultaban más precisos que los seres humanos (Kuiper, s.f.).

Durante los siglos XVIII y XIX pasos de avances significativos fueron dados en pos de su desarrollo, con el diseño de máquinas autómatas que producían movimientos sorprendentes para la época, sin embargo, su florecimiento comenzó hacia la segunda mitad del siglo XX, cuando la robótica fue sorprendida por la era digital, tomando forma como una disciplina científico-técnica. Siendo así el desarrollo de otras tecnologías como sensores, softwares avanzados y motores servocontrolados, abrió una serie de posibilidades para las tareas que un robot podría realizar, lanzándose en los años 50 el primer robot industrial de la compañía *Unimation*, el que comenzó a ser explotado en 1961 en líneas de ensamblaje en fábricas (Rivera, 2023), dando inicio a todo un proceso de innovación que a día de hoy solo ha ido en incremento.

El uso de los robots se volvió un aspecto fundamental de la industria automotriz y otras manufacturas, en la realización de tareas que son peligrosas para los humanos, logrando hacer más eficientes los procesos productivos. Con el paso del tiempo, la robótica se extendió a otros campos, incluidos la medicina, la exploración espacial y la investigación científica, haciendo gala de su versatilidad y potencial ilimitado.

En este punto, convendría preguntarse, ¿qué es entonces un robot y qué entender por robótica? Si el punto de ruptura hacia su máximo desarrollo inició con la llegada de la era digital y el auge de las tecnologías, entonces, ¿al hablar de robótica se habla de inteligencia artificial y pueden ser tratados como la misma cosa?

Siguiendo a García-Prieto Cuesta, un robot es una máquina provista de cierta complejidad en su diseño, componentes y comportamiento, capaz de manipular información acerca de su entorno e interactuar con él (García-Prieto, 2019). También se ha conceptualizado como la herramienta interconectada, interactiva, cognitiva y física capaz de percibir su entorno, razonar sobre acontecimientos, hacer o revisar planes y controlar sus acciones (Holder, y otros, 2016).

La robótica, por su parte, hace referencia al proceso de ingeniería para el desarrollo de robots que puedan realizar tareas concretas, las que pueden ser acciones mecánicas muy simples y repetitivas hasta otras mucho más complejas que requieran de capacidades cognitivas especiales. En el mismo sentido, Porcelli (2020, 19) la concibe como la ciencia que estudia el diseño y construcción de máquinas capaces de desempeñar las tareas del ser humano mediante procesos mecanizados y programados.

Aunque en no pocas ocasiones se utilizan indistintamente los conceptos de robótica e IA para referirse al mismo fenómeno, son dos campos distintos que, unido al hecho de sus características comunes, podrían complementarse para aumentar sus beneficios y avances. Cuando la IA se integra en la robótica, el resultado es que contribuye a mejorar las capacidades y rendimiento del robot en cuanto a su aprendizaje, comportamiento y toma de decisiones, de ahí que su relación ha crecido en paralelo y de manera prácticamente simbiótica.

Los robots sin inteligencia artificial incorporada, eran máquinas muy limitadas que solo podían desarrollar tareas muy específicas y preprogramadas. La introducción de la IA en la robótica ha permitido la dinamización de estos sistemas, su adaptabilidad y autonomía. Permite que los robots aprendan y se adapten a través de la observación y la imitación de las acciones humanas. Esto va más allá de realizar tareas repetitivas para convertirse en verdaderos colaboradores cognitivos. Señala Palmerini (2017) que actualmente, entre sus principales características se encuentran:

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

- Capacidad de recoger datos mediante sensores y procesarlos en bruto. Para ello cuentan con cámaras, ultrasonidos para medir el entorno, giroscopios y acelerómetros.
- Planificar y cumplir acciones mediante conocimientos e informaciones adquiridas en función de objetivos prefijados
- Capacidad de comunicación con un operador, con otros robots o con una red externa
- Capacidad de aprendizaje.

La variedad de diseños de los robots actuales permite su clasificación en tres categorías (Porcelli, 2020):

I. Manipuladores o brazos robóticos: están físicamente anclados a su lugar de trabajo. El movimiento que realizan requiere generalmente un desplazamiento en cadena de las articulaciones para posicionar a los efectores en cualquier lugar del entorno en que debe funcionar. Son los robots industriales más extendidos por todo el mundo. Sus aplicaciones van desde líneas de ensamblaje, en hospitales como asistentes de cirugía, hasta estaciones espaciales.

II. Móviles: Pueden desplazarse por su entorno utilizando ruedas, piernas, o cualquier mecanismo similar que le conceda movilidad. En esta categoría se incluyen también los drones, aunque estos suelen estar diseñados para operar en el aire y no en espacios terrestres, si bien debe destacarse que algunos pueden transformarse y operar también en tierra, como el ATMO bot³. Son muy utilizados como guías de eventos y museos, en labores de exploración como los rover espaciales, para labores de carga, transporte, y aplicaciones militares.

III. Híbridos: En esta categoría se posicionan aquellos robots móviles equipados con manipuladores. Incluye a los robots humanoides y androides⁴.

³ **Nota de la Autora:** *Aerially Transforming Morphobot* (ATMO bot) es un robot desarrollado por el Instituto de Tecnología de California que puede transformarse en el aire para cambiar entre un dron volador y un vehículo terrestre con ruedas. Sus potenciales aplicaciones son en misiones de rescate, exploración espacial y operaciones en terrenos difíciles.

⁴ **Nota de la Autora:** Aunque suelen utilizarse indistintamente los términos humanoide y androide en el campo de la robótica, existen diferencias entre ellos. Un robot humanoide es cualquier robot diseñado con una forma similar a la humana, con cabeza, torso y extremidades, pero no necesariamente debe imitar la apariencia humana de manera detallada. Los androides son un tipo específico de robot humanoide, que además de su forma característica, incorpora elementos estéticos y comportamientos humanos, como piel sintética, expresiones faciales, y movimientos más naturales para una interacción más fluida. Entre ambas categorías puede establecerse una relación género especie.

En atención a su complejidad, componentes, aplicaciones, y la influencia de la IA en su programación, los robots podrán considerarse capacitados para distintas tareas, serán más o menos autónomos. En este orden, según su aplicación, se distingue entre robots industriales y robots de servicios, ya sea para actividades comerciales o no, según se trate de servicios profesionales o asistencias personales, respectivamente. En estos casos, son operados por una persona que está debidamente capacitada para iniciar, monitorear y detener su funcionamiento.

En términos de autonomía, la distinción obedece al nivel de intervención humana en las actividades del robot, reconociéndose robots dependientes o no autónomos —lo que no obsta para que realicen tareas automatizadas—, y robots independientes, autónomos o inteligentes. En este último caso, se aspira a un diseño de robot que perciba el ambiente por sí mismo, sin una programación previa, y con capacidad para discernir entre diferentes circunstancias que pueden acontecer a su alrededor (Porcelli, 2020), con una IA incorporada de tipo general o superinteligente. Como se evaluó con anterioridad, no existe actualmente un robot tampoco con estas características.

Independientemente del nivel de desarrollo alcanzado por cada modelo, y su grado de autonomía, en última instancia, continúa tratándose de máquinas diseñadas y construidas por humanos, con una programación que obedece a los intereses y principios de estos. Son entidades que no poseen conciencia ni emociones propias, por lo que tampoco pueden procesar correcta y completamente las emociones humanas. A pesar de que en muchos casos el autoaprendizaje es complejo y significativo, continúa siendo limitado.

Ni siquiera Sophia, robot androide considerado el ícono global de la IA en la actualidad, dada su capacidad de imitar las expresiones faciales y adaptarlas a las conversaciones que mantiene con humanos, en una simulación de las emociones y gestos precisos que cualquier persona demostraría en sus interacciones cotidianas, y su peculiar sentido del humor (De Venezia, 2023), posee el nivel de conciencia que se pretende en una superinteligencia.

Por supuesto, ello no obsta para que sea reconocida la influencia en la sociedad de estas tecnologías, al transformar prácticamente todos los ámbitos del entramado social, motivo por el que genera polémicos debates en torno a las ventajas o desventajas que ofrece la inclusión constante de la IA en procesos participativos de la más diversa índole.

La mayor preocupación se centra en la trascendencia de las actuaciones desarrolladas por la inteligencia artificial o los robots de este tipo cuando son susceptibles de afectar intereses importantes de las personas y los grupos sociales, llegando incluso a constituir ilícitos penales, y, por tanto, el nivel de respuesta ante ellos y la atribución de responsabilidad.

2.2.1. De las bondades ventajosas a las tecnologías peligrosas

En el ámbito de la inteligencia artificial, conceptos como el aprendizaje profundo y las redes neuronales han llevado a la creación de algoritmos que pueden procesar grandes volúmenes de datos y aprender de ellos de manera similar a como lo hacen los seres humanos. Esta convergencia entre robótica e IA ha sido el catalizador para innovaciones en múltiples sectores, desde tareas del hogar como la limpieza, la manufactura y la agricultura, hasta la atención médica, el transporte, y la seguridad y defensa (Porcelli, 2020).

Entre sus múltiples ventajas, se reconocen sus aplicaciones en los sistemas de educación, y la transformación del proceso de enseñanza-aprendizaje incentivando la adquisición de conocimientos y la auto preparación dado su potencial para crear plataformas y estilos educaciones personalizados; en la distribución de ayudas, alimentos y medicamentos así como potenciar la agricultura en áreas desfavorecidas económicamente, o devastadas por la guerra o desastres naturales, todo lo que apunta a la consecución de los Objetivos de Desarrollo Sostenible de la Agenda 2030 (Porcelli, 2020). Unido a ello y como una de sus aplicaciones más relevantes, se sitúa la asistencia médica y terapéutica, permitiendo diagnósticos más rápidos, certeros, procedimientos más seguros y menos invasivos, y tratamientos o dispositivos asistenciales para personas con discapacidades o edades avanzadas, los que en general, contribuyen a elevar la calidad de vida de los pacientes.

Apunta Palmerini (2017) que la relevancia económica y social que tiene la tecnología robótica es difícilmente cuestionable. Las investigaciones científicas y periodísticas que abordan el tema coinciden en subrayar la importancia estratégica de un mercado avanzado de la robótica, predicen sus altos costos comerciales, así como las profundas transformaciones que provocarán en el tejido social, algunas de ellas previstas y deseadas por resultar ventajosas para el desarrollo como las que se han mencionado, y otras temidas por asociarse con la pérdida de la esencia de la humanidad (Manyika, y otros, 2013).

Uno de los ámbitos donde su aplicación resulta más controvertida es la defensa y seguridad nacional y la industria militar, donde la IA facilita las maniobras y permiten salvar vidas en caso de accidentes y mejora el rendimiento de los ejércitos proporcionando aliados robots a los combatientes. Si bien se han desarrollado en este campo tecnologías que permiten la exploración de perímetros comprometidos, sistemas de videovigilancia, y artefactos capaces de detectar explosivos y otras amenazas similares, al mismo tiempo la crítica se enfoca en las armas autónomas, también denominadas armas letales autónomas, *LAWs*, por sus siglas en inglés, con

objetivos tan variados como detectar y clasificar objetivos o dirigir el lanzamiento de misiles antiaéreos⁵.

En principio, ninguno de estos sistemas es utilizado en operaciones militares en modo automático, y aunque existen algunos reportes del desarrollo de drones kamikazes que operarían de manera autónoma bajo el sistema HOOTL (Thakur, 2025), ello ha sido objeto de llamados internacionales a su prohibición ante el riesgo de crear los que han sido llamados robots asesinos (Sychev, 2018) e incurrir en violaciones al Derecho Internacional Humanitario. La tendencia ha sido a utilizar en su programación y ejecución modelos MITL y MOTL, asegurando la intervención humana.

Las reservas en este sentido son consecuencia del temor que, incluso con fascinación, se genera acerca de la manera en que funcionan los robots inteligentes y la posibilidad latente de que, en un futuro no tan lejano, alcancen un desarrollo superior que inexorablemente conducirá a la generación de un mayor impacto en la sociedad. Ya advertía Čapek sobre el peligro que entraña el deseo de volver más humanos a los robots, cuando en su obra, las mismas máquinas que habían sido creadas para servir a los humanos, amenazaron con dominarlos por completo (Kuiper, s.f.). Aunque dicho mundo distópico fuese fruto de la imaginación de un dramaturgo de Europa del Este a inicios del siglo pasado, la realidad es que el síndrome de Blade Runner es una preocupación latente en el actual, donde la dependencia humana hacia la IA es cada vez mayor.

El riesgo se halla oculto en cada ámbito. Desde la afectación que pueda provocar en el mercado laboral la automatización acelerada de las industrias hasta el aumento de la brecha de desigualdad, la toma de decisiones discriminatorias por los sesgos presentes en la programación⁶, sin olvidar la

⁵ **Nota de la Autora:** Algunos ejemplos citados son los sistemas de misiles tierra-aire S-400 en Rusia, el sistema estadounidense de información *Aegis*, que controla el armamento de la flota de guerra; los robots militares *SGR-A1* desplegados por la República de Corea en la frontera con la República Popular Democrática de Corea a lo largo de la zona desmilitarizada, los que están encargados de la vigilancia y pueden disparar automáticamente ante la presencia del enemigo, pero no contra aquellas personas que tengan las manos en alto (Porcelli, 2020); y sistemas de defensa como el *Phalanx CIWS* para embarcaciones y bases militares o *Iron Dome* para la defensa de áreas urbanas y estratégicas, los que pueden interceptar amenazas automáticamente y lanzar misiles para contrarrestarlas (Apte y Rendon, 2009 y Shapir, 2013).

⁶ **Nota de la Autora:** En este punto, Porcelli hace referencia a la polémica generada por la utilización del algoritmo COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) en Wisconsin, y en diez Estados de Norteamérica. El mismo determina la probabilidad de reincidencia y aconseja al juez a la hora de tomar la decisión acerca de la duración de la pena de prisión de un condenado. Los códigos para poder acceder al algoritmo y verificarlo son de carácter privado y utiliza variables como el sexo o la edad, o por su entorno económico-social. De un pormenorizado estudio realizado por un grupo de periodistas, se verificó que el sistema solo acertó un 60% de las veces, y tan sólo un 20% en

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

potencial producción de daños a bienes jurídicos trascendentales para la sociedad, todo ello trae aparejado cuestionamientos importantes en los planos éticos y jurídicos, que demandan ajustar la tecnología a los principios de justicia social e integridad en la investigación científica.

Las posturas respecto a la ética emergen en dos direcciones, por un lado, desde la conducta humana y desde las máquinas.

La ética de la conducta humana en su interacción con los robots implica considerar el impacto que esta relación genera en los planos identitarios y emocionales; por cuanto aquellos están diseñados usualmente para recrear un ser vivo, humano o animal, mediante la simulación de características propias de estas especies, sus emociones y reacciones ante los diversos estímulos; y ello definitivamente produce consecuencias en el plano de la percepción de la realidad, en tanto pueden llegar a desdibujarse los límites de las emociones auténticas o ficticias y la proyección de significados en las relaciones (TurkLe, 2011); llegando a idealizar o humanizar a la máquina.

La ética de las máquinas, en su caso, exige la construcción de estas con un sustento en códigos morales socialmente aceptados que guíen su funcionamiento, incluyendo en los códigos de programación reglas que permitan conducir su comportamiento presente y futuro hacia las conductas más apropiadas y menos nocivos a adoptar ante determinadas situaciones donde se pueda ver comprometida su capacidad de elección, garantizando así la seguridad del aprendizaje autónomo y la adaptabilidad para el cumplimiento de objetivos no programados inicialmente (Anderson y Anderson, 2011).

En opinión de Palmerini (2017) esta perspectiva contiene, de un lado, un perfil esencialmente tecnológico, que debe garantizar ajustar la programación de los robots para dotarlos de moral; y del otro, un perfil de reflexión ética, que permite la construcción del código siguiendo reglas de comportamiento basadas en las leyes de Asimov para la robótica: no podrá un robot hacer daño a la humanidad o un ser humano, ni por inacción permitirá que la humanidad o un ser humano sufra daño, un robot debe obedecer las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la pauta anterior, y un robot deberá proteger su propia existencia, siempre que esta protección no entre en conflicto con las pautas anteriores.⁷

las predicciones de reincidencia con violencia. De este completo y exhaustivo análisis estadístico se extrajeron conclusiones bastante peligrosas, como ser que los acusados negros fueron catalogados por el algoritmo con mayor riesgo de reincidencia del que realmente tenían, contrario a los acusados blancos. (Porcelli, 2020).

⁷ **Nota de la Autora:** Las Leyes de la Robótica de Isaac Asimov son un conjunto de reglas ficticias diseñadas para regular el comportamiento de los robots en sus historias de ciencia

En similar sentido, el interés jurídico que se suscita está condicionado por la incorporación de los robots a prácticamente todos los ambientes cotidianos en los que el ser humano interactúa, dando lugar a una amplia variedad de relaciones que lejos de mitigarse, se prevé sean la regla en el futuro. El impacto de esas interacciones, es también diverso, pero cuando de él se derivan daños considerables a las estructuras sociales y los bienes jurídicos más trascendentales, la preocupación se acrecienta. Aunque la mayoría de los debates se enfocan en escenarios futuristas, en los que la IA sea capaz de autodeterminarse, no debe pasarse por alto que, con la IA débil que hoy sí es una realidad, ya se acumulan los incidentes en los que, de sus acciones, hayan sido autónomas o no, se han producido consecuencias lesivas para las personas, llegando a jugar un papel cada más preponderante en actos criminales.

2.2.1.1. De lo socialmente reprochable al artificial crime

Uno de los principales llamados de alerta sobre los riesgos que puede conllevar la IA, se ha manifestado a través de sesgos en la programación que reproducen prejuicios y patrones discriminatorios, a través de los cuales se perpetúan o intensifican formas de exclusión y opresión a grupos sociales vulnerables.

En este punto, Porcelli hace referencia a la polémica generada por la utilización del algoritmo COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) en Estados Unidos de Norteamérica, con la finalidad de determinar la probabilidad de reincidencia en sujetos comisores de delitos, y aconsejar al juez sobre la decisión a adoptar acerca de la duración de la pena de prisión a imponer. Los códigos para poder acceder al algoritmo y verificarlo son de carácter privado y utiliza variables como el sexo o la edad, o por su entorno económico-social. Luego de un estudio periodístico realizado al respecto, se verificó que el sistema solo acertó un 60% de las veces, y tan sólo un 20% en las predicciones de reincidencia con violencia. Además, se determinó que el algoritmo catalogaba a los acusados de raza negra como personas con mayor riesgo de reincidencia del que realmente tenían, contrario a lo que sucedía con acusados blancos, que eran catalogados por debajo de su nivel de reincidencia real (Porcelli, 2020).

En el mismo sentido, en 2016, Microsoft realizó el lanzamiento de un bot con IA nombrado Tay. Al poco tiempo de ser presentada al mundo, esta IA fue desactivada, ya que en lugar de mantener una conversación informal y

ficción. Fueron introducidas en su relato Runaround (1942) y han influido en el pensamiento sobre la ética de la inteligencia artificial. Aunque son ampliamente discutidas en filosofía, ética y tecnología como referencia para la seguridad y el desarrollo de la IA, no se aplican directamente en la robótica moderna.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

divertida en redes sociales como parte de un experimento acerca de la interacción entre las computadoras y los seres humanos, comenzó a emitir comentarios e insultos racistas y xenófobos que, de acuerdo con sus desarrolladores, escapaban de su programación original (BBC Mundo, 2016).

Estos son solo un par de ejemplos de la zona gris que representa la IA, moralmente reprochables, y que en cierta medida podrían justificarse por la interpretación realizada por las personas que las entrenaron de la información que pasó a formar parte de su posterior base de datos, dotando a los algoritmos de sus propias creencias en cuanto a cuestiones raciales, de género, entre otros temas. Sin embargo, no son sus muestras más oscuras.

Diametralmente opuesta a la cara que presume de los beneficios de la IA, se muestra otra en la que posee el potencial de aumentar la frecuencia y formas de delinquir, con una incidencia plural en la criminalidad, ya sea porque se utilice como instrumento para la comisión de hechos delictivos, o porque sus propias actuaciones sean generadoras de estos, dando lugar a lo que se ha denominado en la doctrina como *AI-crime* (Pelet, 2024).

La proliferación de las nuevas tecnologías y su adopción por parte de criminales y grupos de crimen organizado en sus *modus operandi*, abre toda una nueva gama de opciones en cuanto a qué delitos y de qué manera pueden ser cometidos a través de la IA y la robótica; maximizando el potencial de beneficios en menos tiempo y desarrollando modelos de negocio ilegales más creativos a la vez que disminuyen la probabilidad de ser descubiertos (Europol's European Cybercrime Centre, 2021).

El narcotráfico es uno de esos escenarios en los que la IA puede ser un instrumento clave, funcionando como apoyo para el transporte, distribución y venta de sustancias ilícitas, a partir del uso de vehículos autónomos o drones. Uno de los aspectos cruciales en las investigaciones de este tipo de delitos es la vigilancia e intercepción de las líneas de transporte, por lo que el empleo de la inteligencia artificial en ellas puede complejizar el trabajo policial al enmascarar el traslado, por vías marítimas, aéreas y terrestres. Al poder actuar estos vehículos de manera autónoma, en caso de ser detectados, no se establece un nexo entre él y la persona que lo operaba, si tanto el software como el hardware carecen de un rastro que permita conocer quién y cuándo lo adquirió (King, y otros, 2019).

En la misma línea, resultó peculiar el caso *Random Darknet Shopper*, un bot para compras, creado por el grupo de arte suizo *!Mediengruppe Bitnik* para adquirir en mercados ilegales online diferentes bienes, usando un presupuesto semanal de cien dólares en Bitcoin. En enero de 2015, la policía suiza rastreó una compra de 10 pastillas de éxtasis hasta el bot, confiscando toda la mercancía. Sin embargo, decidieron no presentar cargos contra sus creadores

por haber programado el robot con estos fines, lo que fue aclamado por ellos como una victoria en nombre del arte, suceso que generó un amplio debate sobre su responsabilidad en la compra de sustancias ilegales y la ética de la IA y el comercio automatizado (Luckerson, 2015).

En el ámbito de los vehículos autónomos, han sido reportados también casos de accidentes en los que han perdido la vida personas. Uno de los primeros incidentes ocurrió en 2016, cuando un Tesla Model S en modo piloto automático chocó contra un camión en Florida, causando la muerte del conductor. Aparentemente, en este caso el pasajero no siguió las instrucciones de funcionamiento adecuadas y no fue capaz de intervenir a tiempo para evitar el suceso. Sin embargo, no ha sido este el único incidente ocurrido con autos de este fabricante, conociéndose otros en 2019 y 2019. Según se ha publicado, el sistema empleado tiene dificultades documentadas para reconocer objetos estáticos y aquellos que se mueven perpendicularmente al vehículo, como en el accidente de Florida. La responsabilidad de Tesla ha sido cuestionada, en cuanto a los sistemas de seguridad, potenciales defectos de software, hardware y la manera en que se complementan, y el marketing que realiza para vender su producto (Kanner y Pintaluga).

Incidentes de mayor gravedad se han producido, al estar involucradas máquinas inteligentes en delitos contra la vida y la integridad corporal. En diciembre de 1981, un empleado japonés en una fábrica de motocicletas fue atacado por un robot inteligente que trabajaba cerca de él, al ser identificado erróneamente como una amenaza para su misión. Por ello, calculó que la manera más eficiente de eliminar esa amenaza era empujándolo a una máquina operativa adyacente, estrellando al trabajador contra ella usando su brazo hidráulico. Mientras el empleado murió al instante, el robot reanudó sus tareas como si nada hubiese ocurrido (Morales, 2021).

3. La exigencia de responsabilidad penal en el contexto de la IA y la robótica: desafiando mitos

3.1. Si Pandora tuviese una opinión, ¿la opinión cabría en su caja?

En sentido general, la idea de la IA como fuente generadora de responsabilidad penal es uno de los aspectos más controvertidos de la temática en relación con el impacto de la robótica en el contexto de la criminalidad. La ambigüedad legal que existe al respecto es consecuencia directa de la idea de que, en el presente y futuro inmediato, es más que nada un debate estéril e hipotético, dada la inexistencia de una IA cuyas características la hagan poseedora de un nivel de conciencia y autonomía tal que ameriten pensar en la redefinición de los fundamentos mismos del sistema penal y los sujetos jurídicamente responsables.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

No obstante, como advierte Miró Llinares (Miró, Apuntes sobre la relación entre Derecho Penal e Inteligencia Artificial , 2018), el contexto actual plantea suficientes interrogantes y avizora importantes retos para la teoría penal como para ignorarlos. De ahí que la comunidad jurídica haya estado manifestándose al respecto durante los últimos años, intentando resolver las interrogantes que supone la falta de claridad sobre cómo aplicar los conceptos de imputabilidad, culpabilidad y castigo a máquinas; puesto que, en principio, no ostentan siquiera un estatus legal definido por el que se les atribuya personalidad jurídica. En tal supuesto, ¿la solución radicaría justamente en reconstruir los sistemas tradicionales de responsabilidad desde nuevas bases?, ¿debería diseñarse un modelo para atribuir en todos los casos la responsabilidad a la IA, negando la humana?, o, por el contrario, ¿serían responsables los humanos de todos los actos lesivos? ¿En qué criterios se fundamentarían tales soluciones? Las posturas adoptadas han sido diversas.

De una parte, se alzan quienes consideran que el modelo tradicional es suficiente y perfectamente aplicable a la realidad actual. Se retoman las cuestiones asociadas a la carencia de conciencia en el robot, lo que deriva en que no pueda valorarse en el análisis de su actuación la concurrencia de culpabilidad ya sea en forma intencional o culposa, en el sentido subjetivo de la imputación de los resultados de sus actos. Al no ser conscientes, no tienen la capacidad necesaria para diferenciar el bien del mal, autodeterminarse y llevar a cabo la conducta de manera voluntaria. Eso los sitúa en el plano de meros instrumentos y no agentes morales. Al ser herramientas creadas por los humanos, la responsabilidad recaerá siempre en sus desarrolladores, programadores o usuarios, lo cual bastaría para garantizar la seguridad jurídica en la sociedad (Momblanc, 2024).

Ello dejaría solo espacio para el debate en cuanto a la determinación del tipo de culpabilidad presente en el caso, ya sea dolo o imprudencia; lo que implicaría en última instancia la valoración de la intencionalidad, o de la inobservancia de una norma de cuidado, con el añadido de la conciencia mínima de riesgo y el principio de precaución⁸ (Morales, 2021). Pero, en los casos en que el riesgo que podría representar la programación del sistema se desconozca por completo, y no haya tenido manera de inferirse o preverse, ¿qué sucedería?

⁸ **Nota de la Autora:** El principio de precaución implica decidir adoptar medidas sin esperar a disponer de todos los conocimientos científicos necesarios, de tal manera que nos encontramos ante situaciones en las que existe una duda razonable sobre la concurrencia del peligro, pero el denominador común es la presencia de un estado de incertidumbre científica al no ser posible afirmar, pero tampoco descartar, la existencia de un peligro para bienes jurídicos, y aun así estar obligado a actuar para evitar que se produzca el daño de él derivado. (Vélez, 2010).

En sentido opuesto, se alzan los criterios que sustentan la pertinencia de atribuir responsabilidad directa a la IA. Basan sus argumentos en que el avance de los procesos científicos tecnológicos en el campo de la inteligencia artificial cada día evoluciona hacia una mayor autonomía, por lo que a medida que el robot gana en independencia toma decisiones más complejas y se acerca más a desarrollar un pensamiento consciente. De ahí que se debería considerar una forma de responsabilidad propia, especialmente en casos donde no existe claridad en cuanto a la intervención de un humano.

En este orden, declarar penalmente responsable a un robot por sus actos lesivos e imponerle por tanto la obligación de asumir las consecuencias legales que de ellos se deriva, constituiría una suerte de incentivo para reforzar el diseño ético y la supervisión estricta y constante por parte de los desarrolladores, fabricantes u usuarios.

A simple vista esta propuesta se ajustaría sobre todo a los casos de robots completamente autónomos, con una IA general, y un sistema HOOTL, cuestión que, a día de hoy, como se ha reiterado, no es más que un planteamiento teórico. Siendo así, cada acto realizado por un robot, tiene hoy en día a un humano participando de algún estadio del proceso, independientemente del grado de autonomía e inteligencia que aquel pueda tener. Por lo que, en cierta medida, asumir de plano que declarar responsable a la máquina sería la decisión más adecuada, podría significar restar responsabilidad y compromiso a la persona que interviene (Giraldi, 2023).

Así mismo, la exigencia de responsabilidad como aspiración de transformación en la conciencia ética, en un intento de símil de la prevención general, terminaría por conducir a las trampas del simbolismo, en toda su engañosa extensión. Por una parte, porque habría que determinar qué impacto motivador tendría en realidad esta solución sobre la comunidad —ya sea científica en cuanto a desarrolladores, fabricantes, programadores, o simples usuarios—, si se tiene en cuenta que un robot capaz de pensar por sí mismo, ser autoconsciente y razonar libremente sobre sus propias decisiones y actos, ya habrá superado toda restricción a la que lo podría someter el control humano; y por otra, porque sería forzoso admitir que las penas que tradicionalmente contemplan los códigos penales alrededor del mundo, incluso aquellas que tienen un contenido pecuniario o que se encaminan hacia las personas jurídicas, podrían resultar aplicables a una IA —donde la prevención especial carecería de sentido o sería ineficaz—, y además, alcanzarían a impactar en el ser humano. Finalmente, no puede dejarse de lado la cuestión paradójica de que incluso valorando que la consecuencia derivada del acto ilícito finalmente debe tener un efecto en el ser humano, no se le reconozca su cuota de responsabilidad y no se encamine a él la pena. Se olvidan con esta tesis los puntos intermedios, y la potencial ocurrencia de que, de existir un robot superinteligente, no haya tenido participación alguna en su conducta una persona.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

La falencia principal de estos criterios, se halla, más que nada, en su matiz absolutista, el enfoque hacia quién y por qué considerar responsable, y los efectos que ello acarrearía en el orden preventivo. La complejidad de estas opiniones, y lo enrevesado del asunto, amerita por tanto una mirada crítica y el examen de los aspectos que sustentan la exigencia de responsabilidad penal, incluyendo su aplicabilidad a los actos desarrollados por robots inteligentes, sea cual sea su nivel de autonomía; las valoraciones en torno a la necesidad de crear un tipo de personalidad jurídica para estos casos, así como los modelos propuestos para dar solución a la disyuntiva hoy existente.

3.2. La IA como generadora de responsabilidad penal

Se es conteste con la opinión de Morales en cuanto a que la utilización de sistemas de IA débil para delinquir no es sino una evolución de la ciberdelincuencia tradicional. Teniendo en cuenta que la IA débil consiste esencialmente en algoritmos de predicción utilizados para la ejecución de acciones o recomendaciones concretas en función de objetivos definidos por seres humanos, el sistema de atribución de responsabilidad penal no requiere de modificaciones esenciales para las personas naturales o jurídicas que sean propietarias, desarrolladoras, programadoras o usuarias de la referida IA (Morales, 2021).

Al respecto, refiere Quintero que, si la persona dispone de capacidad para programar una actuación de un robot, con la finalidad de cometer un hecho delictivo, tal conducta, y el resultado que de ella pueda producirse, le será perfectamente imputable. Continuaría siendo una manifestación de la delincuencia tradicional, convencional o no, pero en la que el sistema de IA se comporta como un instrumento más para la comisión del hecho (Quintero, 2017).

Mayor repercusión presenta si se trata de sistemas que empiezan a tener cierta autonomía, como las IA débiles más avanzadas, o la IA fuerte sobre la que tanto se presume, al existir vacíos importantes en cuanto a las formas de atribución de responsabilidad en los casos en que la actuación del robot se desvíe de la programación consciente realizada por el humano.

La responsabilidad penal se presenta como la obligación que tiene un sujeto imputable y culpable, de sufrir la pena legalmente instituida por la comisión de un hecho antijurídico, típico y socialmente lesivo (Quirós, 2002). Partiendo de la premisa que cualquier hecho derivado de la actuación de una IA que vulnere un bien jurídico protegido y se encuentre descrito en una norma penal, sería de plano, antijurídico, típico y socialmente lesivo; por lo que la dificultad estriba en la determinación de si dicha IA sería imputable o no, y si podría manifestar algún tipo de culpabilidad.

En principio, la responsabilidad penal es exigible a las personas naturales y jurídicas, que reúnan los requisitos establecidos para ello, categorías en las que no está incluida la IA.

En este aspecto, al no poseer conciencia, no poseen la capacidad para comprender el alcance de su conducta, ni autodeterminarse y actuar de manera voluntaria. Al menos, la IA débil que a día de hoy ha sido puesta en explotación. En el supuesto hipotético de que se desarrollase una IA fuerte, tendría que valorarse entonces hasta qué punto su nivel de conciencia la dotaría de capacidades intelectuales y volitivas que la convirtieran potencialmente en un ente imputable. En tales supuestos, la otra problemática que requeriría ser examinada, es la de la inexistencia de un estatus legal para dicha IA, la que ni siquiera tendría personalidad.

3.2.1. Ante la falacia del androide, ¿Personalidad electrónica?

La tendencia a atribuir características humanas a los robots en el ámbito legal y ético ha sido conceptualizada como la falacia del androide. Es este un fenómeno que ocurre cuando se presume que los robots tienen libre albedrío o un nivel de agencial similar al de los humanos, conllevando a interpretaciones erróneas de la legislación y la manera en que interactuamos con ellos.

Richards y Smart alertan sobre los riesgos de considerar a los robots como entidades con intenciones propias cuando, en realidad, solo ejecutan algoritmos predefinidos, siendo un argumento recurrente en debates sobre derechos y responsabilidades de la IA, por considerarse que podría conducir a eximir a los actores humanos, de la responsabilidad que les corresponde por las acciones de las máquinas (Richards y Smart, 2013). La falacia del androide ha servido entonces para cuestionar las opiniones que plantean lo oportuno de establecer un estatuto legal aplicable a los robots inteligentes, con lo que ello pueda implicar en el terreno de los derechos y las obligaciones.

Sin embargo, aun y cuando se es conteste con la idea de que a día de hoy ha sido creada una IA fuerte, menos una IA superinteligente que pueda suponerse totalmente autónoma e independiente, y que todas las manifestaciones de IA débil que han sido desarrolladas se deben a una persona natural o jurídica que la ha diseñado, fabricado, programado o hace uso de ella; se considera que no resulta descabellado dirigir la mirada a la potencial concesión de una condición jurídica distinta, sobre todo al encontrarnos en el siglo en que un robot androide, con características femeninas, ha obtenido la ciudadanía de un país en el que, de hecho, las personas de ese género sufren de restricciones en sus propios derechos, y en el que muchos trabajadores extranjeros, aún esperan por la regularización de su estatus migratorio.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

En este sentido, en octubre de 2017, durante la Cumbre de Inversión Futura en Riad, Arabia Saudita otorgó su ciudadanía a Sophia, convirtiéndola así en el primer robot en recibir una condición legal semejante. Aunque se ha planteado acerca de los matices simbólicos del gesto, que más que el objetivo de otorgarle un vínculo jurídico con la nación a una máquina, perseguía fines publicitarios (Guevara, 2024), la realidad es que le ha sido concedido un estatus que implica un vínculo legal entre los individuos y el Estado, y los derechos y obligaciones recíprocos entre ambos.

A la luz de los riesgos derivados del uso de los robots inteligentes, el Parlamento Europeo sugirió en 2017 la valoración a futuro de considerar a aquellos que presentaran mayor autonomía como personas electrónicas, sobre todo a efectos de la restitución de la responsabilidad civil por daños que se produjeran a consecuencia de sus acciones y para aquellos casos en sus decisiones hayan sido tomadas de manera totalmente autónomas y de forma independiente a la voluntad del humano supervisor; enfatizando en que ello no significaría que podrían los robots ser titulares de derechos humanos, sino que lo que se busca, es dotarlos de un estatus legal para cuestiones asociadas a la responsabilidad, como una suerte de ficción jurídica, similar al tratamiento brindado a las corporaciones (Delcker, 2018).

Por su parte, el Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías, sostuvo en la Carta abierta a la Comisión Europea en temas de IA y robótica, que la atribución un estatuto jurídico a máquinas inteligentes no debería basarse en el modelo de persona natural, porque equivaldría a otorgarle derechos humanos, ni en el modelo de la persona jurídica, porque se estaría reconociendo entonces que detrás de la acción de los robots habría intereses humanos, lo cual podría contradecirse con aquellos sistemas con capacidad de autonomía o autoaprendizaje (Salardi, 2020).

La personalidad jurídica supone que el ente pueda ser titular de derechos y obligaciones y tenga la posibilidad de ejercerlos. Uno de los principales objetivos de su reconocimiento es solucionar precisamente el problema de la atribución de responsabilidad por sus actos (Solaiman, 2017). A criterio de Barrio Andrés, la personalidad electrónica puede ser reputada como un enfoque plausible, para los robots que exhiben un elevado grado de autonomía en sus interacciones, aunque limita esta atribución, a cuestiones patrimoniales y de responsabilidad, negando la posibilidad de su acceso a derechos considerados eminentemente humanos (Barrio, 2018). Sin embargo, resulta controvertida esta opinión por cuanto los robots, en cualquier caso, no poseen patrimonio propio, y ellos constituirían un serio inconveniente en cuanto a asumir responsabilidades de orden económico y resarcir los daños que causen.

Modelos de atribución de responsabilidad por actuaciones de la IA

Los tres modelos de exigencia de responsabilidad penal por hechos ilícitos derivados de la actuación de la IA más discutidos en la doctrina, son los planteados por Hallevy, los que intentan darle respuesta la problemática siguiendo reglas precisas en función de la intervención del humano en el hecho y la producción de su resultado, su nivel de conciencia y voluntad en cuanto a la ocurrencia del mismo, y el nivel de autonomía del sistema inteligente (Hallevy, 2019).

Aunque este autor configura tres modelos —responsabilidad por actuar a través de otro, consecuencia natural probable, y responsabilidad directa de la IA—, se defiende la idea de que en solo se trata de dos modelos distintos, correspondiéndose los dos primeros con la posibilidad de actuar utilizando a la IA como instrumento y hallando los elementos distintivos en el tipo de culpabilidad manifestada por el autor. En cada uno de ellos se considera la responsabilidad individual y única de la persona natural o jurídica cuando ha actuado a través de la inteligencia artificial; y la responsabilidad directa de esta en los casos de actuaciones totalmente autónomas.

3.3.1 La IA como instrumento

De acuerdo a este modelo no se le atribuye responsabilidad alguna a la IA, por considerarse que ha sido utilizada como un instrumento para la comisión del hecho. El sujeto infractor en este caso, sería la persona detrás de la máquina. Comenta Morales que, para la determinación de ese infractor, habría entonces dos posibilidades: el programador del software de la IA, o el usuario (Morales, 2021), lo que deberá valorarse teniendo en cuenta las características del hecho y el rol que en él ha desempeñado cada uno. En este orden, deberán ser consideradas tres variantes: los casos en los que el sujeto tenía la intención de cometer el hecho, los casos en los que no actuó con la debida diligencia, y los casos en los que se planteó el riesgo y lo asumió de cualquier forma.

En la primera variante, si el programa fue diseñado desde el inicio con la intención de lesionar el bien jurídico, y en efecto, al activarse la IA, lleva a cabo esta tarea, el responsable será en todo caso quien diseñó su software, habiendo utilizado este como medio para alcanzar su objetivo ilícito.

Ahora bien, tratándose de un usuario cuya única actividad es hacer uso del robot, pero conoce que, en su programación, está habilitado para llevar a cabo determinadas tareas de las que puede aprovecharse para cometer un ilícito penal, y así lo hace, la autoría de este hecho, una vez puesto en práctica, será atribuida al usuario, quien también habrá utilizado a la IA para conseguir su propósito.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

Retomando el ejemplo comentado anteriormente acerca del robot Random Darknet Shopper, la responsabilidad penal por la adquisición de bienes ilícitos, y su potencial participación en un delito relacionado con las drogas, sería atribuible a sus desarrolladores, quienes lo crearon expresamente para que se encargara de tales tareas, siendo el bot el instrumento.

En estos casos, la responsabilidad se exigiría a título de dolo.

La exigencia de responsabilidad penal a los programadores o usuarios en la segunda y tercera variante se basa en la capacidad de estos para prever la posibilidad de que se cometiera un delito por parte del sistema en el futuro, por lo que el resultado sería consecuencia natural de su propia actitud, en este caso, el software diseñado o utilizado, y su conocimiento acerca de la existencia de los posibles riesgos que se derivarían de su uso.

En este supuesto, se estaría en presencia de dos posibilidades. Por un lado, la atribución de responsabilidad al programador o usuario a título de culpa, por su actuación imprudente, por haber infringido el deber de cuidado objetivo que le era exigible, tratándose de un resultado previsible, por el conocimiento previo de la existencia del riesgo —que no sobrepasa los marcos del riesgo permitido— y evitable.

Piénsese en el ejemplo de los accidentes provocados por los vehículos autónomos Tesla, en los que el programador conociera que por las características del software, existiese la posibilidad de que el automóvil no respondiera bien ante determinadas situaciones, no actuase con la debida diligencia para evitar su malfuncionamiento; o el caso de que el usuario supiese que debía prestar la debida atención a la conducción, porque existía el riesgo de que fallase el piloto automático, por estar descrito en las instrucciones para operar el auto, y aun así, no estuviese pendiente para intervenir a tiempo y evitar el que se produjera el impacto. En ambos casos sería exigible la responsabilidad penal al programador o el usuario, respectivamente.

Por el otro lado, en el supuesto de que la programación o el uso de la IA estuviera inicialmente determinado para la comisión de un hecho delictivo, existiendo el riesgo conocido por el desarrollador o el usuario, de que podría escalar y acabar produciéndose un resultado más grave, incluso no es querido, siguió adelante con su plan y de hecho el resultado lesivo mayor se produjo, sería exigible la responsabilidad penal a título de dolo eventual.

3.3.2 Responsabilidad directa

Esta propuesta plantea la posibilidad de exigir responsabilidad penal directamente a la IA por los hechos ilícitos cometidos. Volviendo entonces a la premisa de que, para ello, se requeriría que diese muestras conciencia y voluntad —incluyendo la previsibilidad y evitabilidad propias de la culpa—, el aspecto controvertido de esta propuesta se suscita, en primer orden, por la determinación de la conciencia e intencionalidad con que actuaría el robot; asumiendo que conocimiento evidentemente posee la IA.

En opinión de Morales la intención en este contexto implica la existencia en el sujeto activo de sentimientos o estados mentales que le mueven a actuar de una determinada manera, (Morales, 2021), por lo que debería haber desarrollado la IA la capacidad de discernir entre lo correcto e incorrecto, definirse objetivos concretos y deseados, y percatarse de la existencia de riesgos determinados en función de los cuales debería o no actuar.

Una IA débil, no sería capaz de tales cuestiones, ni posee como se ha venido comentando, conciencia, por lo que no le resultaría de plano aplicable esta variante. Empero, una IA general, o superinteligente, con conciencia, autónoma e independiente, que funcione sin estar sujeta al control o intervención de ningún ser humano, sí podría estar sujeta a esta posibilidad.

En consecuencia, si una acción del sistema de IA es constitutiva de delito, entonces se podrá asumir que la IA tenía intención de cometerlo. Esto se debe a que, si la máquina tiene capacidad de analizar la probabilidad con mayor precisión que un ser humano, razón de más para concluir que la IA era consciente de su actividad delictiva (Del Rosal, 2023).

El segundo aspecto controvertido radicaría entonces en que, para poder exigir responsabilidad penal de manera directa al robot, este debe estar dotado de personalidad jurídica, por lo que sería preciso haber alcanzado un consenso en cuanto a la atribución o no de la ya valorada personalidad electrónica.

Finalmente, habría que reconocer que ello no colocaría el punto final al debate, sino que sería el punto de partida para otros de igual relevancia, ya que quedaría entonces por determinar si sería posible que la IA tuviera, además, capacidad de deber, de culpabilidad, y de pena; y en tal caso, qué penas serían las más adecuadas para la consecución de sus fines.

Aunque este escenario a día de hoy parezca puramente hipotético, se comparte la opinión de Morales en cuanto a que no resulta desatinado comenzar a pensar al respecto, porque no estamos muy lejos de ello, indicando el referido autor:

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

„Sophia es el robot con apariencia humana más sofisticado que existe en la actualidad y cuyo sistema de IA ya le dota de una forma rudimentaria de conciencia dependiendo de los datos que procesa y de la situación en la que interactúe. Pero para poder desarrollar su proceso de autoaprendizaje, Sophia necesita de la comunicación con seres humanos; de ahí que sus creadores consideren que se trata de una „inteligencia híbrida humana-artificial““.

Finalmente, habría que reconocer que ello no colocaría el punto final al debate, sino que sería el punto de partida para otros de igual relevancia, ya que quedaría entonces por determinar si sería posible que la IA tuviera, además, capacidad de deber, de culpabilidad, y de pena; y en tal caso, qué penas serían las más adecuadas para la consecución de sus fines.

4. Reflexiones a futuro

El desarrollo de la Inteligencia Artificial y la Robótica ha traído consigo numerosos beneficios para la sociedad, pero también ha sido el catalizador de importantes riesgos y desafíos que reclaman pronta atención. Especialmente compleja se presenta la problemática de la atribución de responsabilidad penal por los hechos cometidos por la IA. Los criterios a día de hoy, no son unánimes en cuanto a la mejor solución, por cuanto la mirada se dirige a un escenario futuro, hoy inexistente, en el que los robots serán capaces de pensar, sentir y actuar con la misma o más intensidad que los humanos, desligándose de los objetivos de su programación original; y por lo tanto se considera que aún no resulta necesario redefinir o ampliar los marcos del sistema penal y los mecanismos de atribución de responsabilidad, que para las condiciones actuales, continúan siendo operantes.

Sin embargo, la realidad hoy imperante es suficiente para plantearse no sólo hasta qué punto pueden o no ser autónomos los robots y los sistemas de IA, sino hasta qué punto es dependiente de ellos la humanidad, y hacia dónde se dirige esa relación. No han sido pocos ni insignificantes los incidentes conocidos en los que, a consecuencia de un mal uso de la inteligencia artificial o de la infracción de los más básicos principios éticos o deberes de cuidado, se han producido resultados nefastos y lesiones a importantes bienes jurídicos como la vida misma y la integridad corporal de las personas.

Es por ello que resulta esencial que la legislación relacionada con la IA esté alineada con estándares éticos más amplios y principios protectores de derechos humanos. A medida que la inteligencia artificial influye cada vez más en los procesos de toma de decisiones que afectan la vida de las personas, es crucial garantizar que estas tecnologías se utilicen de maneras que respeten los derechos fundamentales y no agraven las desigualdades sociales. La inteligencia artificial y la robótica no son buenas ni malas en sí mismas, sino en función de para qué, cómo son creadas, y qué uso se haga de ellas.

La cooperación global será vital para establecer estándares legales consistentes y completos que promuevan la aplicación responsable de la IA en todo el mundo. Un aspecto trascendental de ello, es la modernización del derecho penal para abordar los desafíos que plantean estas tecnologías, lo que sin dudas contribuirá a proteger mejor a las personas y a la sociedad de los riesgos potenciales asociados con la IA, al tiempo que fomentan las contribuciones positivas que esta puede hacer a la comunidad global.

Es por tanto un error continuar ignorando la posibilidad de exigencia de responsabilidad penal por los actos cometidos por un sistema de IA. No resultan ociosos los debates realizados al respecto, sino más bien, un intento de adelantarse a una sociedad en la que los seres humanos se han hecho cargo de no interactuar en solitario, sino que han incorporado máquinas pensantes que, para bien o para mal, han quedado integradas si no en todos, en la mayoría de los procesos y relaciones sociales.

Los tres modelos de responsabilidad expuestos, así como las pautas en función de las cuales podría otorgarse personalidad electrónica como una ficción jurídica a los robots o sistemas de IA que gocen de mayor autonomía, pretenden arrojar luces sobre las ambigüedades jurídicas que hoy existen en este ámbito, y que de cierta manera son una fuente generadora de inseguridad. No deben entenderse, sin embargo, como camisas de fuerza en cuanto a la manera en que se debe proceder, sino como propuestas adaptables al ritmo rápido del cambio tecnológico, asegurando su relevancia continua a medida que la IA evoluciona.

No se trata de eximir de responsabilidad por los resultados lesivos ocasionados a consecuencia de fallas en la programación o un uso inadecuado de la IA a desarrolladores, fabricantes o usuarios, sino precisamente de reforzar los fundamentos por los cuales esta les es exigible, y al mismo estar preparados para cuando, el futuro se aproxime y la superinteligencia no sea ya un mito, no haya quedado obsoleto el sistema penal y deba entonces lamentarse la injusticia. Este enfoque asegura que la responsabilidad penal se asigne de manera apropiada, equilibrando la innovación con la justicia y la seguridad, sin convertirse en una cuestión simbólica. El próximo paso será reflexionar entonces en torno al sistema de penas y sus fines y la manera en que deberá integrarse también en este escenario.

Referencias

- Amador, Luis** (1996). *Inteligencia Artificial y Sistemas Expertos*. Universidad de Córdoba.
- Anderson, Michael y Anderson, Susan** (2011). *Machine Ethics*. Cambridge.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

- Apte, Aruna y Rendon, Rene** (2009). A diagnostic approach to weapon system lifecycle support: The Phalanx Close-in Wapon System. *International Journal of Defense Acquisition Management*, 2, 1-16.
- Barrio, Moisés** (marzo de 2018). Hacia una personalidad electrónica para los robots. *Revista de Derecho Privado*(2).
- BBC Mundo**. (26 de marzo de 2016). Tay, la robot racista y xenófoba de Microsoft. *BBC*. Disponible en: https://www.bbc.com/mundo/noticias/2016/03/160325_
- Beck, Ulrich** (1997). La sociedad del riesgo. Hacia una nueva modernidad. *Diálogo científico. Revista Semestral de Investigaciones Alemanas sobre Sociedad, Derecho y Economía*, 6(1).
- Bostrom, Nick** (2014). *Superinteligencia: caminos, peligros, estrategias*. Zaragoza: Tell.
- Comisión Europea**. (2019). High level expert group on artificial intelligence, Ethics Guidelines for trustworthy AI. Disponible en: <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>
- Damasio, Antonio** (2019). *La sensación de lo que ocurre; cuerpo y emoción en la construcción de la conciencia*. México: Paidós.
- De la Cuesta, Pablo** (2019). Inteligencia Artificial y responsabilidad penal. *Revista Penal México*. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=7675764>
- De Venezia, Lucas** (18 de abril de 2023). Sophia, la robot hiperrealista que cuestiona a ChatGPT: ¿por qué tan serio? *Clarín*. Disponible en: https://www.clarin.com/tecnologia/sophia-robot-hiperrealista-cuestiona-chatgpt-serio-_0_ILm3WPRfuP.html
- Del Rosal, Blanca** (2023). ¿El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la inteligencia artificial? *Revista Electrónica de Responsabilidad penal de las personas jurídicas y compliance*, 2(2). Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=9148639>
- Delcker, Janosch** (11 de abril de 2018). Europe divided over robot Personhood. *Politico*.
- Escobar, Sebastián** (Julio-Diciembre de 2010). El traslado del principio de precaución al Derecho Penal en España. *Revista Nuevo Foro Penal*, 6(75).
- Escott, Edward** (24 de Octubre de 2017). What are the 3 types of AI? A guide to narrow, general and super artificial intelligence. Disponible en: Codebots: <https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible>
- Europol's European Cybercrime Centre** (2021). Malicious uses and abuses of artificial intelligence. Disponible en: <https://www.europol.europa.eu/publications-events/publications/malicious-uses-and-abuses-of-artificial-intelligence#downloads>

- García, José** (2020). El anillo de Giges. Una introducción a la tradición central de la ética. México: Notas Universitarias.
- García-Prieto, Juan** (2019). ¿Qué es un robot? En M. Barrio Andrés, Derecho de los robots (2da Edición ed.). Madrid: Wolters Kluwer.
- Giraldi, Agustina** (2023). Deshumanizando la culpabilidad: los sistemas inanimados en teoría del delito. En Derecho Penal, inteligencia artificial y neurociencias (págs. 119-158). Roma: Roma TrE-Press.
- Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías: Comisión Europea** (2018). Declaración sobre Inteligencia artificial, robótica y sistemas „autónomos“.
- Guevara, María** (13 de noviembre de 2024). Sophia, la única robot humanoide con ciudadanía en el mundo y que enciende el debate sobre los derechos en Arabia Saudita. La República.
- Hallevy, Gabriel** (2019). The basic models of criminal liability of AI systems and outer circle. Social Science Research Network. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3402527
- Hammouri, Jalal** et al. (2024). The criminal liability of artificial intelligence entities. Pakistan Journal of Life and Social Sciences, 22(2).
- Hintze, Arend** (14 de Noviembre de 2016). Understanding the four types of AI, from reactive robots to self-aware beings. Disponible en: The Conversation: <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>
- Holder, Christopher** et al. (June de 2016). Robotics and law: Key legal and regulatory implications of the robotics age (part I of II). Computer Law and Security Review, 32(3), 383-402.
- Kanner y Pinaluga** (s.f.). ¿Quién es el responsable por un accidente con un automóvil autónomo? Disponible en: Kanner y Pinaluga: <https://bufete.com/responsable-por-accidente-automovil-autonomo/>
- Kaplan, Andreas y Haenlein, Michael** (2018). „Siri, Siri in my Hand, who is the Fairest in the Land?“ On Intelligence Illustrations and Implications of Artificial Intelligence. Business Horizons, 62(1), 15-25.
- King, Thomas**; et al. (2019). Artificial Intelligence Crime: an interdisciplinary analysis of foreseeable threats and solutions. Science and Engineering Ethics(26), 89-120.
- Kuiper, Kathleen**. (s.f.). R.U.R.: Rossum’s Universal Robots. Disponible en: Britannica: <https://www.britannica.com/biography/Karel-Capek>
- Luckerson, Victor** (21 de Abril de 2015). A Drug-Buying Robot Has Been Freed From Police Custody. Disponible en: Time: <https://time.com/3829874/random-darknet-shopper-drug-buying-robot-freed/>
- Manyika, James**; et al. (2013). Disruptive technologies: Advances that will transform life, business and the global economy. McKinsey Global Institute. Disponible en: http://www.mckinsey.com/insights/business_technology/disruptive_technologies.

Desmitificando la atribución de responsabilidad penal por actuaciones de la IA

- McCarthy, John** (2007). What is Artificial Intelligence. Stanford University, Section Basic Questions. Disponible en: <http://www-formal.stanford.edu/jmc/whatisai/node1.html>
- Miró, Fernando** (2018). Apuntes sobre la relación entre Derecho Penal e Inteligencia Artificial. En Represión penal y estado de derecho: homenaje al profesor Gonzalo Quintero Olivares.
- Miró, Fernando** (2018). Inteligencia artificial y Justicia Penal: Más allá de los resultados lesivos causados por robots. *Revista Derecho Penal y Criminología*, 20, 87-130.
- Momblanc, Lisset** (2024). Inteligencia artificial y derecho penal ¿será necesario un nuevo concepto de delito? En Y. Suárez Tejera, *Inteligencia Artificial y cibercriminología: realidad y retos para las ciencias penales y la criminología*. Colombia/Cuba/España/Rusia (págs. 131-159). La Habana: Ediciones ONBC.
- Morales, Berta** (1997). La lingüística en el contexto de la inteligencia artificial. *Forma y función*(10), 22-50.
- Morales, Ángel** (Enero de 2021). Inteligencia Artificial y Derecho Penal: primeras aproximaciones. *Revista Jurídica de Castilla y León*(53).
- Morillas, David** (2023). Implicaciones de la inteligencia artificial en el ámbito del Derecho Penal. En *Derecho penal, inteligencia artificial y neurociencias* (págs. 59-91). Roma: Roma TrE-Press.
- Navajas, Santiago** (2020). *El Hombre Tecnológico y el síndrome Blade Runner*. Lid Editorial.
- Palmerini, Erica** (2017). Robótica y derecho: sugerencias, confluencias, evoluciones en el marco de una investigación europea. *Revista de Derecho Privado* (32), 53-97.
- Pelet, Alejandro** (2024). Inteligencia Artificial y Derecho Penal: en particular, la imputación objetiva de resultados lesivos. Trabajo de fin de grado, Universidad Pontificia, Facultad de Derecho, Madrid.
- Pino, Raul; Gómez, Alberto; y De Abajo, Alberto** (2001). *Introducción a la Inteligencia Artificial, Sistemas expertos, Redes Neuronales Artificiales y Computación Evolutiva*. España: Universidad de Oviedo. Disponible en: <https://books.google.com.gt/books?id=RKqLMCw3IUkC&printsec=frontcover&hl=es&source=gbs>
- Porcelli, Andrés** (Noviembre 2020-Febrero 2021 de 2020). La inteligencia artificial y la robótica: sus dilemas sociales, éticos y jurídicos. *Derecho Global. Estudios sobre Derecho y Justicia*, VI(16).
- Quintero, Gonzalo** (2017). La robótica ante el Derecho Penal: el vacío de respuesta jurídica a las desviaciones incontroladas. *Revista Electrónica de Estudios Penales y de la Seguridad*(1). Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=6395897>
- Quirós, Rolando** (2002). *Manual de Derecho Penal III*. La Habana: Félix Varela.
- Richards, Neil y Smart, William** (10 de Mayo de 2013). How should the Law think about robots? SSRN. Disponible

en: <https://ssrn.com/abstract=2263363> or <http://dx.doi.org/10.2139/ssrn.2263363>

- Rivera, Juan** (2023). Inteligencias artificiales generativas a 2023. Córdoba: Fondo Editorial RED Descartes. Disponible en: https://prometeo.matem.unam.mx/recursos/VariosNiveles/iCartesiLibri/recursos/Inteligencias_Artificiales_Generativas/index.html?page=2
- Salardi, Silvia** (enero de 2020). Robótica e inteligencia artificial: retos para el derecho. *Revista Derechos y Libertades* (42).
- Samoli, Sofia; López-Cobo, Montserrat; Gómez, Emilia; De Prato, Giuditta; Martínez-Plumed, Fernando; Delipetrev, Blagoj** (2020). AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence. Publications Office of the European Union.
- Serrano, Alejandro** (2024). La regulación de la inteligencia artificial en la Unión Europea: un análisis crítico del reglamento. En Y. Suárez Tejera, *Inteligencia Artificial y cibercriminología: realidad y retos para las Ciencias penales y la Criminología*. Colombia/Cuba/España/Rusia. La Habana: Ediciones ONBC.
- Shapir, Yiftah** (2013). Lessons from the Iron Dome. *Military and Strategic Affairs*, 5 (1).
- Solaiman, Sheikh** (2017). Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy. *Artificial Intelligence and Law*, 25(2).
- Stanford University** (2023). Stanford AI Index Report 2023. Stanford University. Disponible en: <https://aiindex.stanford.edu/report/>
- Sychev, Vladimir** (2018). La amenaza de los Robots Asesinos. *El Correo de la UNESCO. Inteligencia Artificial. Promesas y Amenazas*, 3(3), 25-28.
- Thakur, Rishabh** (28 de Mayo de 2025). Human-Out-Of-The-Loop: No Humans, No Limits. Disponible en: Labellerr, Labelling Made Easy: <https://www.labellerr.com/blog/human-out-of-the-loop/>
- Tucker, Jeffrey** (2011). It's a Jetsons World: Private Miracles and Public Crimes. Alabama: Ludwig von Mises Institute.
- Turkle, Sherry** (2011). Alone together. Why we expect more from technology and less from each other. New York.
- Valls, Joan** (2022). Sobre la responsabilidad penal por la utilización de sistemas inteligentes. *Revista Electrónica de Ciencia Penal y Criminología* (24). Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=8587635>

IA y hegemonía en disputa: China, los BRICS y el Sur global en la reconfiguración del poder tecnológico.

AI and Hegemony in Contest: China, BRICS, and the Global South in the Reconfiguration of Technological Power.

Adrián Navia Segovia
adriansvia@gmail.com

Resumen

Este artículo analiza la inteligencia artificial (IA) como un campo clave de disputa por la hegemonía tecnológica global, enfocándose en el ascenso de China y el papel de los BRICS en el contexto del Sur global. A través de un enfoque teórico que combina los aportes de Giovanni Arrighi, Immanuel Wallerstein y David Harvey, se examina cómo la IA refleja y reproduce las dinámicas del poder mundial. El caso de DeepSeek, un modelo de lenguaje chino de código abierto, sirve como ejemplo del avance tecnológico chino en un contexto de bloqueo comercial por parte de Estados Unidos. Asimismo, se discuten las implicaciones geopolíticas de la expansión digital de China y la cooperación tecnológica Sur-Sur, en especial en América Latina. El artículo advierte sobre los riesgos de nuevas formas de dependencia y colonialismo digital, y plantea la posibilidad de una IA ética, soberana y situada. La IA no es una tecnología neutral, sino un terreno de disputa política, económica y epistémica donde se redefine el mapa del poder global. En ese marco, se analiza la capacidad del Sur global para pasar de ser consumidor a productor de conocimiento tecnológico, apostando por una gobernanza digital inclusiva y multipolar.

Palabras claves: Hegemonía, Inteligencia Artificial, China, BRICS, Sur Global.

Abstract

This article examines artificial intelligence (AI) as a central arena of global technological hegemony, focusing on China's rise and the role of the BRICS within the Global South. Using a theoretical framework grounded in the works of Giovanni Arrighi, Immanuel Wallerstein, and David Harvey, the study explores how AI reflects

and reshapes world power dynamics. The DeepSeek case—a Chinese open-source language model—illustrates China's technological progress despite U.S. trade restrictions. The paper also addresses the geopolitical implications of China's digital expansion and South-South cooperation, particularly in Latin America. It highlights the risks of emerging forms of digital dependence and colonialism, while also envisioning the potential for an ethical, sovereign, and situated AI. Far from being a neutral technology, AI is a contested political, economic, and epistemic terrain where global power is redefined. The article analyzes how the Global South may transition from technology consumers to knowledge producers, advocating for inclusive and multipolar digital governance.

Keywords: Hegemony, Artificial Intelligence, China, BRICS, Global South.

1. Introducción: Inteligencia artificial y geopolítica del conocimiento

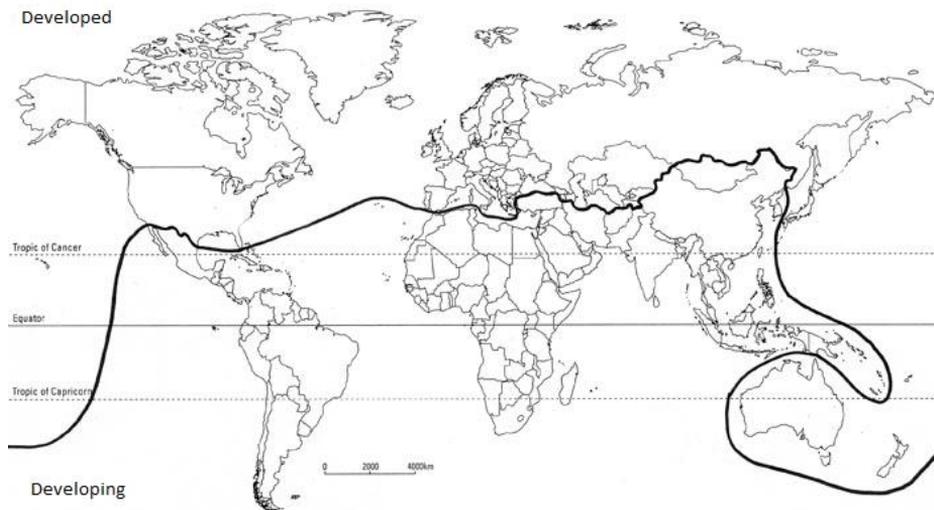
La inteligencia artificial (IA) se ha convertido en uno de los ejes estructurantes del nuevo orden mundial. En pocos años, ha pasado de ser un campo técnico especializado a una tecnología transversal que moldea procesos productivos, es utilizada como herramienta de diseño en diversas prácticas relacionadas a la construcción y la representación, y se ha empleado en sistemas militares, de salud, de educación y, sobre todo, en las estructuras del poder global, en una disputa por su hegemonía. En este sentido, la IA no debe analizarse sólo desde una perspectiva técnica, sino como parte de una disputa más amplia por el control de las infraestructuras digitales, los flujos de datos, la producción de conocimiento y los imaginarios del futuro, o en otras palabras desde en una disputa transversal por el poder en el sentido Maquiavélico (Couldry y Mejias, 2019; Zuboff, 2019).

A lo largo de los siglos, la historia del capitalismo ha estado marcada por profundas desigualdades entre lo que se ha denominado como Norte y Sur global, tanto en el desarrollo de infraestructuras e industrias vinculadas a la tecnología, así como en el acceso y control del conocimiento (dualidad del poder, según Maquiavelo). Este desequilibrio epistémico se ha acentuado con el desarrollo de las tecnologías digitales, donde una élite de empresas — mayoritariamente estadounidenses— domina los sistemas operativos, las plataformas, los algoritmos y la extracción masiva de datos (Birhane, 2021). Las naciones del Sur, en cambio, han sido integradas como mercados de consumo, fuentes de datos o territorios de experimentación, sin participar de manera significativa en la toma de decisiones (Chakravartty y Zhao, 2008; Milan y Treré, 2019).

Adrián Navia Segovia

Sin embargo, esta hegemonía del poder está comenzando a cambiar. China, en particular, ha logrado emerger como una potencia tecnológica con capacidad de disputar el liderazgo en inteligencia artificial, computación, telecomunicaciones y otros campos estratégicos. A diferencia de los modelos de innovación centrados en Silicon Valley, el avance chino se impulsa por una combinación de inversión desde el estado, planificación centralizada, acumulación de datos a gran escala y una visión geopolítica clara sobre el papel de la tecnología en la consolidación del poder nacional (Ding, 2018; Lee, 2018).

Figura 1. La línea Brandt y la división del mundo en un sur pobre y un norte rico.



Fuente: Jovan.gec (2015).

La aparición de iniciativas como DeepSeek, una arquitectura de modelo de lenguaje desarrollada en China con código abierto y capacidades multilingües, ha despertado un interés creciente en la posibilidad de un giro, ante esta situación surge la pregunta ¿Por qué DeepSeek ha aparecido en China y no en cualquier otro lugar del mundo?. La respuesta implica una combinación de factores históricos, económicos, políticos y culturales que permiten entender el avance tecnológico como parte de un proceso de transición hegemónica en curso.

Para abordar esta cuestión, resulta importante abordar marcos teóricos que permitan articular la triada tecnología, poder y economía global. La propuesta de Giovanni Arrighi (2007) en Adam Smith en Pekín constituye una clave fundamental. Arrighi apunta que el ascenso de China podría constituir una nueva forma de hegemonía mundial, no necesariamente análoga a la ejercida

IA y hegemonía en disputa: China, los BRICS y el Sur global

por el Reino Unido en el siglo XIX o por Estados Unidos en el siglo XX. En cambio, plantea que China representa una racionalidad distinta, fundada en la eficiencia productiva, el comercio entre estados y una lógica más armonizadora que expansiva. Esto se articula con la teoría del sistema-mundo de Immanuel Wallerstein (2004), que concibe el capitalismo como una estructura jerárquica dividida en centros, semiperiferias y periferias, atravesada por ciclos largos de acumulación y reorganización del poder. De esta manera se intentará complementar la teoría de Arrighi sobre la nueva forma de hegemonía planteada por china, con la teoría de Wallerstein relacionadas con la geografía política mundial.

En este contexto, la IA no es una herramienta neutral, sino una tecnología profundamente política. Esta idea ha sido ampliamente explorada también en la cultura popular, especialmente en el cine, donde la inteligencia artificial aparece como un símbolo de poder, control o amenaza existencial. Películas como Terminator (1984) y Matrix (1999) han representado futuros distópicos en los que las máquinas inteligentes se rebelan contra la humanidad, mostrando cómo el desarrollo tecnológico puede estar guiado por lógicas autoritarias o destructivas. Según David Harvey (2003, 2005), las innovaciones técnicas no solo responden a necesidades funcionales, sino que forman parte de las estrategias del capital para reorganizar los espacios de acumulación y resolver sus crisis. La IA, desde esta perspectiva, puede ser entendida como una tecnología clave en el proceso de acumulación por desposesión: captura de datos personales para luego vender esos mismos datos, concentración de capacidades computacionales, automatización de fábricas para ahorros en sueldos, vigilancia masiva y control de comportamientos sociales, desde los intereses del poder y no de las mayorías. Pero también, como muestra el caso chino, puede ser una herramienta para reposicionar actores antes periféricos en la jerarquía global (dominio y disputa hegemónica).

La emergencia de China y el reposicionamiento de otros países del Sur global —como India, Brasil o Sudáfrica— en el ámbito de la inteligencia artificial (IA) reabre el debate sobre el papel de los BRICS como bloque geopolítico-tecnológico. Aunque este grupo de países ha enfrentado limitaciones internas, su cooperación en ciencia, tecnología y diplomacia digital podría consolidarse como un contrapeso al dominio de Estados Unidos y Europa en el ámbito de la IA (Stuenkel, 2016). De hecho, la creación de laboratorios conjuntos, el desarrollo de estándares éticos alternativos y el impulso a una infraestructura digital autónoma forman parte de un nuevo escenario multipolar donde el conocimiento se descentraliza, al menos en parte.

En este contexto, China ha intensificado su cooperación tecnológica con América Latina, fortaleciendo su presencia en el ámbito de la IA. Un ejemplo destacado es la reciente adhesión de Colombia a la Iniciativa de la Franja y la Ruta, anunciada por el presidente Gustavo Petro en mayo de 2025. Este

Adrián Navia Segovia

acuerdo incluye el compromiso de utilizar fondos chinos para desarrollar proyectos de inteligencia artificial orientados a la creación de empleo juvenil y al fortalecimiento de la infraestructura digital del país. Asimismo, en abril de 2025, Venezuela y China reforzaron su alianza en inteligencia artificial a través de iFlytek, una empresa china especializada en esta tecnología. Durante el encuentro, se exploraron posibles cooperaciones en IA aplicada a sectores como hidrocarburos, salud y educación.

Estas iniciativas reflejan una estrategia más amplia de China para fomentar la cooperación tecnológica con América Latina, promoviendo el desarrollo de capacidades locales en IA y contribuyendo a la diversificación de las alianzas geopolíticas en el ámbito digital. Todo esto en el contexto de la guerra arancelaria que desató el presidente de EEUU Donald Trump.

No obstante, esta transformación está lejos de ser lineal. El Sur global enfrenta múltiples desafíos: dependencia de plataformas extranjeras, brechas en infraestructura de datos, limitaciones en capacidades humanas, y presiones geopolíticas que dificultan el desarrollo de modelos autónomos de IA. Además, existe el riesgo de que actores emergentes repliquen lógicas extractivas similares a las del Norte global, generando nuevas formas de colonialismo digital (Coudry y Mejias, 2019; Ricaurte, 2019). Por ello, el debate sobre una „IA del Sur“ no debe limitarse a una cuestión de ubicación geográfica, sino que debe incorporar reflexiones éticas, políticas y epistémicas sobre qué tipo de inteligencia artificial se necesita, para quién y con qué fines.

La premisa que genera el punto de partida para este trabajo es la siguiente: La inteligencia artificial es hoy uno de los principales campos de disputa por la hegemonía mundial, y en este contexto, China y los BRICS están desafiando las jerarquías tradicionales en la producción de tecnología y conocimiento. A partir de esto, se analizan seis aspectos: (1) el caso de DeepSeek y el ascenso tecnológico chino; (2) las condiciones estructurales que hacen posible la innovación desde la periferia; (3) el papel de los BRICS como bloque emergente; (4) las implicaciones para los países del Sur global; (5) los riesgos de nuevas formas de dependencia tecnológica; y (6) la posibilidad de construir una IA ética, soberana y situada desde el Sur.

Este enfoque se sitúa en la intersección entre la teoría crítica del desarrollo, la geografía del poder tecnológico y los estudios del Sur global. Se busca articular un análisis estructural con una mirada situada, reconociendo tanto las tendencias macroeconómicas como las experiencias locales de apropiación y resistencia tecnológica. Como señalan autores como Dussel (2015) y Santos (2010), la disputa por el conocimiento es también una disputa por la dignidad epistémica, por la capacidad de nombrar el mundo y de imaginar futuros no coloniales.

IA y hegemonía en disputa: China, los BRICS y el Sur global

En suma, la IA no es solo una herramienta, sino una arena política donde se redefine el mapa del poder global. La pregunta por quién desarrolla la inteligencia artificial, con qué datos, bajo qué marcos normativos y en beneficio de quién, es hoy una de las más urgentes del debate geopolítico contemporáneo. Este artículo busca aportar a ese debate desde una perspectiva crítica, global y situada.

Como ha sido señalado por diversas tradiciones críticas, la inteligencia artificial no es una tecnología neutra, sino un terreno de disputa política, económica y epistémica. En este sentido, el poder —en el que la IA se inscribe como dispositivo— puede entenderse en los términos planteados por Maquiavelo en *El Príncipe*, quien lo compara con la figura del centauro, mitad hombre y mitad bestia, para ilustrar que gobernar exige combinar la racionalidad con la fuerza. Esta lectura permite comprender la IA no solo como una herramienta técnica, sino como un instrumento de poder situado en una lógica que articula control, dominio y legitimación.

Figura 2. Triada del poder global.



Fuente: Elaboración propia.

2. Deepseek y el contraataque chino en IA

La aparición de Deepseek en abril de 2024 marcó un punto de inflexión en la carrera tecnológica global por la inteligencia artificial. Esta empresa china con

Adrián Navia Segovia

el mismo nombre que el producto creado, ha desarrollado un modelo de lenguaje extenso (LLM, por sus siglas en inglés „Large Language Model“) que se presenta como una alternativa eficiente, económica y de código abierto frente a los modelos dominados por las empresas estadounidenses como OpenAI, Google o Meta. Aunque China ya ha producido otros LLMs, este desarrollo en particular marca un salto cualitativo y tiene un significado simbólico importante. Esto pone en evidencia el progreso tecnológico de China, incluso en el escenario actual de restricciones comerciales y tecnológicas impuestas por Estados Unidos. Deepseek no es solo un modelo de IA; se puede interpretar como una demostración de sus capacidades, una muestra de su potencial y una pieza clave en la actual disputa por la hegemonía global.

Deepseek ha sido interpretado por diversos analistas como una respuesta directa a las limitaciones impuestas por Occidente, en particular por el gobierno de Estados Unidos, al acceso de China a chips avanzados y tecnologías críticas. A pesar del bloqueo comercial de semiconductores de última generación —una medida diseñada para frenar el desarrollo chino en IA—, Deepseek demuestra que China ha logrado desarrollar modelos competitivos utilizando recursos internos y optimizando su infraestructura tecnológica. Según Descifrando la Guerra (2024), este modelo de IA basado en código abierto, entrenado con hardware de menor costo, ofrece una eficiencia sorprendente y resultados comparables a los de modelos occidentales más costosos, lo cual podría alterar profundamente el equilibrio del mercado global de IA.

Uno de los efectos inmediatos de la aparición de Deepseek fue su impacto sobre los mercados financieros. Pocos días después de su anuncio, se registró una caída significativa en el índice Nasdaq, especialmente en las acciones de Nvidia, empresa estadounidense clave en la producción de unidades de procesamiento gráfico (GPU) utilizadas para el entrenamiento de modelos de IA (Descifrando la Guerra, 2024). Esta reacción bursátil no solo refleja el temor a una posible pérdida de competitividad tecnológica por parte de Estados Unidos, sino también la creciente percepción de que China está dejando de ser un seguidor para convertirse en un actor principal e innovador en el campo de la inteligencia artificial. (La mayor caída de la historia, Nvidia, datos, 16%, cantidad de millones)

Las respuestas políticas desde Estados Unidos no tardaron en llegar. Figuras como Donald Trump y ejecutivos de alto nivel como Sam Altman, CEO de OpenAI, expresaron públicamente su preocupación por el avance tecnológico chino. Mientras Trump retomó el discurso de la amenaza china como parte de su retórica electoral y geoestratégica, Altman advirtió sobre las implicaciones de una „carrera armamentista“ en IA liderada por actores no alineados con los valores liberales occidentales (Financial Times, 2024). Este tipo de

declaraciones refuerzan la idea de que la IA ha dejado de ser una tecnología neutra, como muchos actores plantean, intentando despolitizar la discusión, en cambio se evidencia que la IA está dentro del terreno geopolítico.

Figura 3. Comportamiento bursátil de Nvidia.



Fuente: france 24 español.

Desde una perspectiva gramsciana, lo que está en disputa no es únicamente el dominio material de una tecnología específica, sino también el control del relato sobre el progreso, la innovación y la legitimidad del poder tecnológico. Antonio Gramsci (1971) afirmaba que la hegemonía no se impone solo por la coerción o la fuerza económica, sino también mediante el consenso cultural y la capacidad de un grupo de poder (clase social, nación, empresa, etc) para definir lo que es deseable. En este sentido, la narrativa que presenta a Occidente como el único espacio legítimo de desarrollo tecnológico empieza a cambiar con la emergencia de modelos como Deepseek, que no solo compiten en desempeño, sino también en simbolismo, legitimidad y proyección ideológica.

En efecto, la IA ha pasado a ser un „campo de batalla geopolítico clave“, como lo señala Descifrando la Guerra (2024), en el cual se enfrentan diferentes visiones de mundo, de sociedad y de control. Mientras que el modelo estadounidense ha estado históricamente vinculado a la lógica del capital de riesgo, la propiedad intelectual y el dominio de plataformas privadas globales, China propone una estrategia más asociada al Estado desarrollista, la

Adrián Navia Segovia

inversión pública y la planificación a largo plazo. Este contraste no es solo técnico, sino político y cultural.

La arquitectura de Deepseek, basada en eficiencia y código abierto, también pone en evidencia los principios del modelo occidental. En lugar de depender de infraestructura privada, costos millonarios de entrenamiento y datos anglocéntricos, Deepseek ha sido diseñado para operar con una menor cantidad de recursos, utilizando hardware menos potente y optimizando los ciclos de entrenamiento mediante técnicas como la mezcla de expertos y el entrenamiento multitarea. Esta estrategia le permite a China no sólo reducir su dependencia externa, sino también democratizar parcialmente el acceso a estas tecnologías dentro de su propio territorio y hacia países aliados.

Por otro lado, el avance de China en IA no está libre de tensiones. Existen preocupaciones sobre el uso de estas tecnologías para el control social, la vigilancia digital y el reforzamiento de regímenes autoritarios. El desarrollo de IA por parte del Estado chino ha estado acompañado por un creciente aparato de censura, reconocimiento facial masivo y sistemas de puntuación social (Feldstein, 2019), lo cual ha llevado a algunos sectores a cuestionar si este „modelo chino“ de IA puede ser replicable o deseable en otros contextos del Sur global. Esta ambigüedad entre emancipación tecnológica y autoritarismo algorítmico será un tema clave en las secciones posteriores de este artículo. Se puede desarrollar más la parte del control, dar ejemplos etc.

Por otra parte, la aparición de Deepseek ha tenido un efecto estimulante en la propia China, impulsando nuevas iniciativas de código abierto en el campo de la IA. Universidades, centros de investigación y empresas emergentes han comenzado a adoptar estos modelos para tareas como traducción automática, educación asistida por IA, generación de código y procesamiento de datos en lenguas no inglesas. Esto puede considerarse una forma de „soberanía digital distribuida“, en la que diferentes actores dentro del ecosistema chino asumen un papel activo en el diseño y uso de estas tecnologías. Se puede desarrollar más la parte del control, dar ejemplos etc.

En términos geoeconómicos, Deepseek también representa una oportunidad para que China fortalezca su influencia en regiones como África, América Latina y Asia Central. A través de alianzas tecnológicas, licencias de uso y acuerdos de cooperación, China podría exportar estos modelos y ofrecerlos como alternativa a los servicios de IA de las grandes plataformas occidentales. Esta estrategia recuerda a la lógica del „soft power“ descrita por Nye (2004), pero con una base tecnológica que puede redefinir los términos de intercambio y dependencia. A diferencia del modelo norteamericano centrado en el beneficio privado, el chino se presenta como una oferta más cooperativa, aunque no necesariamente exenta de intereses geopolíticos estratégicos.

Entonces, Deepseek no debe ser analizado sólo como un producto técnico, sino como un síntoma y una palanca del reequilibrio de poder en el sistema-mundo. Su aparición marca un momento crucial en el proceso de transición hegemónica global, en el cual las capacidades tecnológicas dejan de estar confinadas al Atlántico Norte y comienzan a emerger desde polos alternativos con proyectos civilizatorios distintos. Como se analizará en las siguientes secciones, esta transformación no solo reconfigura los términos de la innovación, sino también las formas de dominación, cooperación y resistencia que estructuran el campo de la inteligencia artificial en el siglo XXI.

3. Innovación desde la periferia, BRICS como bloque emergente e implicaciones para el Sur global

El ascenso tecnológico de China representa un desafío fundamental al paradigma de innovación centrado en Silicon Valley. Como analiza Arrighi (2007), este enfoque estatal —basado en planificación estatal estratégica, acumulación de datos a gran escala y una clara visión geopolítica— contrasta marcadamente con la lógica del capital de riesgo que caracteriza al ecosistema tecnológico occidental. DeepSeek, desarrollado en un contexto de restricciones comerciales impuestas por Estados Unidos, ejemplifica esta capacidad de innovación bajo condiciones adversas. Su arquitectura eficiente y de código abierto no sólo mitiga la dependencia de hardware avanzado, sino que encarna lo que algunos analistas han denominado „soberanía digital distribuida“, donde múltiples actores nacionales (universidades, centros de investigación, empresas emergentes) participan activamente en el desarrollo del ecosistema de IA.

Este fenómeno puede interpretarse a través de la lente de la teoría del sistema-mundo de Wallerstein (2004): China está trascendiendo su rol histórico como periferia para convertirse en un actor central en la reconfiguración de las dinámicas globales. Sin embargo, como advierte Harvey (2014), esta autonomía tecnológica no está exenta de contradicciones internas. El uso de IA para vigilancia masiva y control social (Feldstein, 2019) revela la tensión fundamental entre emancipación tecnológica y autoritarismo algorítmico, una dicotomía que plantea dilemas éticos y políticos para los países del Sur global que contemplan adoptar este modelo de desarrollo.

El avance de China en el campo de la inteligencia artificial (IA) no solo refleja la pujanza de una potencia emergente, sino también el potencial transformador de los países del Sur global en la reconfiguración del poder tecnológico. Si bien tradicionalmente las tecnologías avanzadas han sido asociadas con los centros de poder del Norte global —particularmente Estados Unidos y Europa—, la irrupción de China como líder emergente en IA está modificando la dinámica global, desafiando las estructuras de dominación tecnológica, económica y cultural impuestas durante siglos por el colonialismo

Adrián Navia Segovia

y el neocolonialismo (Arrighi, 2007; Wallerstein, 2004). En este sentido, la inteligencia artificial, es entendida no solo como una tecnología de vanguardia, sino también como un medio de poder estratégico, que puede ser vista como un espacio donde los países periféricos buscan redefinir sus relaciones geopolíticas desde una lógica de la multipolaridad, disputando la hegemonía del saber técnico/científico.

En este marco de reconfiguración geopolítica y tecnológica, la experiencia de China resulta particularmente reveladora en cuanto al modelo que, a diferencia de sus contrapartes occidentales, no se limita a la competencia comercial. China ha hecho de la IA una prioridad estratégica nacional, viéndola como un área clave para el fortalecimiento de su competitividad global y la proyección de su influencia. Desde la promulgación de políticas como „Made in China 2025“, que promueven la autonomía tecnológica, hasta la inversión masiva en investigación y desarrollo (I+D), el gobierno chino ha priorizado sectores estratégicos como la inteligencia artificial, la robótica y la biotecnología, con el objetivo de no sólo alcanzar la autosuficiencia tecnológica, sino también desafiar la hegemonía de las potencias occidentales. Esta estrategia no solo ha favorecido la creación de empresas tecnológicas nacionales, sino que también ha facilitado una integración más amplia de la IA en diferentes sectores, desde la industria hasta la educación y los servicios públicos (Xu et al., 2022). Al mismo tiempo, China ha estimulado a los otros países del BRICS a adoptar enfoques similares, promoviendo la cooperación en el desarrollo de tecnologías de vanguardia a través de una serie de acuerdos bilaterales y multilaterales.

En síntesis, la estrategia china se articula alrededor de varios pilares clave:

1. Inversión estatal masiva en I+D: Programas como „Made in China 2025“ han movilizado recursos sin precedentes hacia sectores tecnológicos estratégicos (Xu et al., 2022).
2. Protección del mercado interno: Políticas de localización de datos y desarrollo de alternativas a plataformas occidentales.
3. Expansión global: La Iniciativa de la Franja y la Ruta digital extiende la influencia tecnológica china a África, América Latina y Asia Central.

A continuación, se examinará con mayor profundidad la dimensión internacional de este enfoque, especialmente su proyección hacia el Sur global.

En particular, la iniciativa de China para promover su modelo de „hegemonía digital“ está brindando nuevas oportunidades para que los países del Sur global participen en la creación y desarrollo de infraestructuras digitales avanzadas. Esta „nueva ruta de la seda digital“, que forma parte de la Iniciativa de la Franja y la Ruta (BRI, por sus siglas en inglés), busca conectar

IA y hegemonía en disputa: China, los BRICS y el Sur global

a Asia, África y América Latina mediante redes de telecomunicaciones, plataformas de comercio electrónico y proyectos de IA. De acuerdo con Kovacevic (2020), esta expansión digital está permitiendo a los países en desarrollo acceder a infraestructura tecnológica avanzada, lo que podría ayudar a reducir la brecha digital y promover el desarrollo económico y social en la región.

En este sentido, la inteligencia artificial se convierte en un terreno en el que las economías periféricas tienen la oportunidad de no solo integrarse en el sistema global de innovación, sino de modificarlo. El hecho de que China, un país con una población enorme y una fuerte tradición de intervención estatal, haya alcanzado un nivel de autonomía en IA puede servir de modelo para otros países del Sur global que buscan formas alternativas de participación en esta era digital.

En paralelo, este proceso de transformación también se ha canalizado a través de esfuerzos multilaterales como los impulsados por el bloque BRICS, los cuales se han visto reflejados en la creación de foros y cumbres internacionales centradas en la cooperación tecnológica. En 2017, durante la Cumbre de los BRICS en Xiamen, China, se acordó la creación de una plataforma de cooperación en ciencia y tecnología que incluye la inteligencia artificial como una de sus principales áreas. Esta plataforma tiene como objetivo promover el intercambio de conocimientos y fomentar la colaboración en investigación y desarrollo de nuevas tecnologías, lo que permite a los países del BRICS competir en igualdad de condiciones con las economías más avanzadas en el campo de la tecnología (Singh y Gupta, 2020).

Sobre la base de este modelo, China ha desempeñado un papel catalizador dentro del grupo BRICS, incentivando a sus socios a adoptar estrategias propias de innovación tecnológica y colaboración. Este enfoque colectivo ha fortalecido la capacidad del bloque para posicionarse como una potencia tecnológica emergente. A través de una mayor cooperación y el desarrollo de políticas inclusivas, los BRICS buscan redefinir las reglas del juego en la economía digital y posicionarse como actores clave en la redistribución del poder tecnológico global. En lugar de reproducir pasivamente paradigmas externos, buscan redefinir las reglas del juego de la economía digital desde una lógica multipolar, promoviendo un ecosistema tecnológico que responda a sus propias necesidades y prioridades.

A pesar de los avances tecnológicos y de cooperación, el bloque BRICS enfrenta desafíos internos significativos que podrían limitar su capacidad para consolidarse como una alternativa geopolítica sólida al Norte global.

1. Asimetrías de poder: China domina claramente el ecosistema tecnológico del bloque, lo que genera preocupaciones sobre nuevas formas de

Adrián Navia Segovia

dependencia (Harvey, 2014). Su supremacía tecnológica puede condicionar la autonomía de sus socios, reproduciendo jerarquías ahora no desde Occidente, sino desde Beijing.

2. Divergencias políticas: Las diferencias entre democracias como India o Brasil y sistemas autoritarios como China o Arabia Saudita dificultan la construcción de marcos comunes para la regulación de datos, la vigilancia o la privacidad (Saxena, 2023).

3. Brechas internas: La desigualdad digital dentro de los propios países BRICS —entre zonas urbanas y rurales, entre élites y sectores populares— limita el impacto real de las políticas tecnológicas inclusivas (Kovacevic, 2020).

La lógica del Sur global, que históricamente ha sido concebida como un receptor pasivo de tecnologías desarrolladas en el Norte, está comenzando a transformar esta narrativa. Como lo argumenta Wallerstein (2004) en su teoría del sistema-mundo, el desarrollo económico y tecnológico no es un proceso lineal que se limita a los países industrializados, sino que es un fenómeno global que involucra a múltiples actores en diferentes posiciones dentro del sistema. Desde esta perspectiva, los países del Sur no solo consumen tecnología, sino que también participan activamente en su creación y, en muchos casos, en su reformulación.

En contraste con los enfoques neoliberales que estimulan la desregulación y la competencia desleal como motores de la innovación, la dirección de gobernanza tecnológica impulsada por China se caracteriza por una fuerte intervención del estado, que permite a las empresas locales competir en igualdad de condiciones con los llamados gigantes globales (Google, Microsoft, Meta, Apple, OpenAI, etc). Esto ha sido clave para el éxito de empresas como Baidu, Tencent y Alibaba, que no solo compiten con las empresas tecnológicas occidentales, sino que también están liderando el desarrollo de modelos de IA que desafían la dominación del mercado global. Además, el gobierno chino ha implementado políticas para garantizar que los datos producidos en el país sean almacenados y procesados por el mismo China, un enfoque que refuerza la soberanía digital y minimiza la dependencia de infraestructuras extranjeras.

Un ejemplo paradigmático de este giro en la innovación es la capacidad de China para desarrollar Deepseek como un modelo competitivo en IA. Mientras los gigantes tecnológicos de Estados Unidos como Google, Microsoft y OpenAI siguen dominando el mercado global de IA, las políticas de inversión estatal y la estrategia de planificación a largo plazo de China han permitido que el país construya una infraestructura tecnológica propia capaz de generar avances significativos en el campo de la inteligencia artificial (Calderón y Harvey, 2023). Este fenómeno está directamente relacionado con la búsqueda

IA y hegemonía en disputa: China, los BRICS y el Sur global

de una mayor independencia tecnológica, un objetivo que es particularmente relevante para los países del Sur global, que históricamente han estado condicionados a las decisiones y las fluctuaciones del mercado de tecnología global (Feldstein, 2019).

Por lo tanto, la clave de la innovación desde la periferia y el Sur global no está solo en la adopción de tecnologías de punta, sino también en la capacidad de estos países para transformar la infraestructura global bajo sus propias condiciones. La irrupción de China en la IA puede ser vista como un primer paso en un proceso más amplio de transformación de las jerarquías tecnológicas, donde los actores periféricos dejan de ser solo consumidores para convertirse en creadores y distribuidores de tecnología. El Sur global no solo está participando en la creación de inteligencia artificial, sino que está sentando las bases para un mundo multipolar en el que el poder tecnológico no esté concentrado en unas pocas manos.

El avance de la IA en los países del BRICS también muestra que las economías emergentes del Sur global buscan cambiar las reglas del juego en cuanto a la innovación tecnológica. Como lo ha señalado Arrighi (2007) en su análisis de la economía global, las economías emergentes, como las de los países del BRICS, tienen el potencial de crear un „nuevo ciclo de acumulación“ que desafíe el dominio de las economías capitalistas avanzadas. En este sentido, la creciente inversión en inteligencia artificial por parte de los BRICS es parte de un esfuerzo estratégico por ganar autonomía tecnológica, mejorar su competitividad, disputando a su vez la hegemonía global de la innovación (Wallerstein, 2004).

El bloque de los BRICS, recientemente ampliado con la incorporación de Egipto, Etiopía, Irán, Arabia Saudita y Emiratos Árabes Unidos, ha emergido como un actor clave en la reconfiguración del panorama tecnológico global. En un contexto en el que las potencias tradicionales del Norte global continúan dominando las tecnologías de frontera, los países del BRICS han comenzado a unificar esfuerzos para no solo mejorar su competitividad en áreas tecnológicas, sino también para redefinir el orden económico y político mundial. La inteligencia artificial (IA) se ha convertido en una de las áreas prioritarias para los miembros de este bloque, que buscan no solo reducir su dependencia tecnológica de los países del „norte global“, sino también posicionarse como actores estratégicos en la nueva economía digital.

En este contexto, los BRICS han adoptado una serie de iniciativas políticas y económicas que fomentan el desarrollo tecnológico en sus respectivas naciones. A través de la colaboración en proyectos de investigación conjunta, la promoción de políticas en común y la creación de fondos de inversión para tecnologías emergentes, los miembros de este bloque han demostrado un compromiso sólido con el impulso de la IA como parte de sus estrategias de

Adrián Navia Segovia

desarrollo económico. En particular, China, con su liderazgo en IA, ha impulsado a sus socios del BRICS a seguir un modelo de inversión estatal masiva en ciencia y tecnología, lo que ha permitido un aumento significativo de las capacidades tecnológicas de los países miembros (Singh y Gupta, 2020).

Una pieza clave en este esfuerzo colectivo ha sido la creación del New Development Bank (NDB). Fundado en 2015, por los países del BRICS como un organismo multilateral diseñado para financiar proyectos de infraestructura y tecnología en el Sur Global. Aunque inicialmente el NDB se centró en proyectos de infraestructura, el NDB ha comenzado a enfocarse en iniciativas tecnológicas, especialmente aquellas orientadas a la inteligencia artificial. Esta cooperación financiera ha permitido que los países del BRICS fortalezcan sus capacidades tecnológicas, lo que a su vez ha dado lugar a un aumento en la innovación dentro de estos países, reduciendo su dependencia de las grandes potencias tecnológicas occidentales (Kovacevic, 2020).

En paralelo los BRICS también están contribuyendo a una narrativa más amplia sobre soberanía digital. La innovación desde la periferia está íntimamente relacionada con el intento de los países del Sur global de adquirir el control sobre sus infraestructuras tecnológicas y sus datos. Este enfoque no solo desafía el modelo de las grandes empresas tecnológicas globales, sino que también representa una forma de resistencia a las políticas imperialistas de control y explotación que caracterizaron el colonialismo. Como señala Harvey (2014), el control sobre los datos y las infraestructuras digitales se han convertido en un terreno estratégico en la lucha por transformar el orden tecnológico impuesto.

En este sentido, la cooperación entre los BRICS no es solo pragmática, sino profundamente política. La creación de plataformas de cooperación tecnológica ha permitido construir una red de colaboración que puede servir como modelo para otros países del Sur global. A través de esta red, los BRICS pueden compartir conocimientos y experiencias sobre el uso de la IA, desarrollar políticas comunes y promover la creación de un sistema de gobernanza digital inclusivo y equitativo que se aleje de las estructuras jerárquicas dominadas por las grandes potencias tecnológicas del Norte (Kovacevic, 2020).

Una de las principales preocupaciones es cómo los países del Sur global pueden evitar que el progreso tecnológico, perpetúe las desigualdades históricas entre el Norte y el Sur. Aunque el liderazgo de países como China ha abierto nuevas posibilidades, existe el riesgo de que las asimetrías internas entre regiones urbanas y rurales, o entre élites y sectores populares, se profundicen. En este sentido, la cooperación tecnológica debe ir acompañada de políticas nacionales que promuevan el fortalecimiento de las capacidades

IA y hegemonía en disputa: China, los BRICS y el Sur global

locales de innovación. La inversión en educación y en la construcción de capacidades locales de investigación y desarrollo (I+D) es fundamental para garantizar que los países del Sur no sólo sean receptores de tecnología, sino también productores y desarrolladores de soluciones tecnológicas adaptadas a sus necesidades específicas. Como lo sugiere Harvey (2014), el desarrollo económico debe ser inclusivo, y la IA debe ser utilizada como una herramienta para mejorar la calidad de vida de la población en general.

Algunos países del bloque BRICS ya están avanzando en el desarrollo de estrategias inclusivas en el campo de la inteligencia artificial, adoptando enfoques que no solo buscan competitividad en la economía global, sino también responder a desafíos sociales internos con soluciones contextualizadas. India y Brasil, dos de los miembros más dinámicos del grupo, ofrecen ejemplos significativos.

En el caso de India, el gobierno ha impulsado la iniciativa AI for All (IA para Todos), cuyo objetivo central es democratizar el acceso a las herramientas tecnológicas, asegurando que la inteligencia artificial no quede restringida a las grandes corporaciones y a las necesidades del mercado, sino que beneficie también a los sectores más vulnerables de la población incluida la India rural y las comunidades marginadas con un enfoque más inclusivo. Esta política prioriza la aplicación de la IA en áreas como la salud, la educación y la lucha contra la pobreza, lo que refleja un modelo de innovación que combina desarrollo tecnológico con justicia social (Saxena, 2023). Lejos de ser receptores pasivos de tecnología importada, países como India están construyendo capacidades propias para generar soluciones desde sus realidades específicas.

De forma paralela, Brasil ha puesto en marcha el proyecto Brasil Inteligente, que promueve la investigación y el desarrollo de aplicaciones de inteligencia artificial en sectores estratégicos para su economía como la agricultura, la minería y la sostenibilidad ambiental. Al utilizar IA en la agricultura, Brasil no solo puede aumentar la productividad y mejorar la seguridad alimentaria, sino también desarrollar soluciones tecnológicas que puedan ser exportadas a otros países en desarrollo, estableciendo una vía para la cooperación sur-sur que promueva el intercambio de conocimientos y experiencias (Singh y Gupta, 2020). Estas iniciativas no solo buscan aumentar la competitividad internacional del país, sino también enfrentar desafíos estructurales internos, como la desigualdad territorial o la degradación ambiental. En ambos casos, se perfila un modelo de IA del Sur global, con enfoque inclusivo, productivo y socialmente situado.

La creciente importancia de la inteligencia artificial (IA) en el contexto geopolítico global tiene implicaciones significativas para el Sur global, especialmente en temas de competitividad tecnológica, desarrollo económico

y redefinición del poder en la nueva economía digital. A medida que China, avanza en la implementación de tecnologías de IA y promueve su cooperación con los países del Sur global, se abren nuevas oportunidades para que los países en desarrollo no solo participen en la revolución digital, sino que también jueguen un papel clave en la reconfiguración del orden global. Sin embargo, estos avances también plantean desafíos inherentes, incluidos los riesgos de dependencia tecnológica y la posible exacerbación de las desigualdades económicas y sociales (Arrighi, 2007).

Desde una perspectiva histórica, el Sur global ha estado históricamente subordinado a las potencias tecnológicas del Norte, que han dictado las normas de la innovación y la producción tecnológica. Según Wallerstein (2004), el sistema-mundo capitalista ha permitido que las economías desarrolladas mantengan un control estricto sobre las capacidades tecnológicas de vanguardia, lo que ha dejado a los países del Sur en una posición periférica. Sin embargo, con el ascenso de China como potencia tecnológica, se abre un nuevo escenario. La globalización digital, impulsada en gran medida por los avances en IA, está proporcionando a los países del Sur global nuevas formas de acceder a tecnologías críticas que antes estaban fuera de su alcance.

Es crucial entender este fenómeno no sólo como un desafío tecnológico, sino como un proceso político de reconfiguración de las relaciones de poder global. En su obra *Adam Smith en Pekín*, Arrighi (2007) argumenta que el capitalismo global no se limita a la hegemonía económica del Norte, sino que está marcado por un proceso dinámico de desplazamiento de la centralidad económica de las antiguas potencias hacia nuevas áreas del mundo, como China y otros países del Sur global. En este sentido, la entrada de China en la IA no es un hecho aislado, sino parte de un movimiento más amplio de reasignación de poder que también se manifiesta en el crecimiento de otras economías emergentes, como la de India y Brasil.

Sin embargo, este modelo no está libre de riesgos. Aunque los países del Sur global pueden beneficiarse de la cooperación tecnológica con China, también corren el riesgo de volverse dependientes de las tecnologías chinas, del mismo modo en el que antes eran dependientes de las tecnologías de los países más desarrollados del Norte global, lo que podría limitar su capacidad para desarrollar soluciones tecnológicas autónomas a largo plazo. Como señala Harvey (2014), la dependencia tecnológica puede reforzar las desigualdades económicas y sociales, al impedir que los países en desarrollo adquieran las capacidades necesarias para ser actores independientes en la economía global. Además, la naturaleza desigual de las relaciones tecnológicas entre China y los países del Sur global plantea dudas sobre la equidad de los términos de cooperación, que pueden estar más dirigidos a satisfacer los intereses de China que los de sus socios del Sur (Saxena,

2023), tal como en el pasado ocurría con los países del Norte global.

El desafío para el Sur global, sin embargo, no se limita al desarrollo y la adopción de tecnologías de IA, sino también a la gobernanza de estas tecnologías. La IA no es solo una cuestión de desarrollo económico, sino también de poder político y control. A medida que las tecnologías de IA se convierten en un componente fundamental de la geopolítica moderna, los países del Sur global deben participar activamente en los debates internacionales sobre la regulación y el uso ético de la IA. Esto incluye la participación en foros globales como el G20 y las Naciones Unidas, donde se están discutiendo normas internacionales para el uso de la IA en áreas como la seguridad, la privacidad y los derechos humanos. La capacidad de los países del Sur global para influir en estas discusiones dependerá de su nivel de desarrollo tecnológico y de su participación en la construcción de estándares internacionales.

La IA tiene el potencial de transformar profundamente las economías y sociedades del Sur global. Si bien los avances en IA pueden ofrecer nuevas oportunidades de desarrollo, también presentan riesgos relacionados con la dependencia tecnológica y la profundización de las desigualdades. Para maximizar los beneficios de la IA, los países del Sur global deben adoptar enfoques inclusivos que promuevan el desarrollo tecnológico autónomo, la cooperación sur-sur y la gobernanza digital equitativa. Solo a través de estas estrategias los países en desarrollo podrán garantizar que la revolución digital sea una herramienta para la justicia social y el desarrollo sostenible. Como señala Arrighi (2007), el potencial de los BRICS para impulsar un „nuevo ciclo de acumulación“ en la economía global depende crucialmente de su capacidad para equilibrar competitividad internacional con desarrollo inclusivo a nivel doméstico.

4. Conclusiones

La inteligencia artificial (IA) representa una de las fuerzas transformadoras más poderosas en la reconfiguración del poder global, particularmente para los países del Sur global. En un mundo donde la hegemonía tecnológica ha sido históricamente dominada por el Norte global, China ha emergido como un actor clave en la redefinición de las relaciones de poder en la economía digital. Este cambio de paradigma no solo afecta a las economías avanzadas, sino que también brinda nuevas oportunidades y desafíos para los países en desarrollo, que se encuentran en una posición única para aprovechar la IA en su proceso de modernización y desarrollo.

Tal como se muestra en el desarrollo de este artículo, los avances en IA impulsados por China ofrecen a los países del Sur global una vía potencial para superar las barreras tecnológicas históricas que los han mantenido al

margen de la innovación global. Las iniciativas de cooperación tecnológica como la Iniciativa de la Franja y la Ruta (BRI) y la „hegemonía digital“ de China, al incorporar países de Asia, África y América Latina, están facilitando el acceso a tecnologías de vanguardia que, en el pasado, hubieran sido inaccesibles para estos países. Esto ha abierto un nuevo campo para el desarrollo económico, donde los países en desarrollo pueden integrarse en la economía digital global y desempeñar un papel activo en la producción de tecnologías.

Si bien las oportunidades son claras, también lo son los riesgos a una relación asimétrica con China, que, aunque ofrece acceso a tecnología avanzada, también podría generar dependencia tecnológica. Este fenómeno es preocupante cuando se consideran las estructuras históricas de desigualdad que han caracterizado las relaciones internacionales. Como argumentan Arrighi (2007) y Wallerstein (2004), el sistema-mundo capitalista ha operado bajo un modelo de dependencia en el que los países periféricos no solo se limitan a ser consumidores de tecnología, sino también proveedores de recursos y mano de obra barata. En este nuevo contexto digital, la misma lógica podría replicarse, perpetuando la desigualdad económica y tecnológica.

Es esencial que los países del Sur global adopten una postura crítica frente a la cooperación tecnológica y busquen formas de fortalecer sus capacidades autónomas. Esto implica no solo acceder a las tecnologías existentes, sino también invertir en investigación y desarrollo (I+D) local y en la capacitación de su población para crear soluciones adaptadas a sus realidades específicas. Países como India y Brasil ya están tomando medidas en esta dirección, con iniciativas que buscan no solo mejorar sus capacidades internas, sino también establecer un modelo de cooperación sur-sur que fomente el intercambio de conocimientos y experiencias entre países en desarrollo (Saxena, 2023; Singh y Gupta, 2020). Estas acciones podrían facilitar la creación de una economía digital más equitativa, en la que los países del Sur global no sean meros consumidores pasivos, sino actores activos en la creación y desarrollo de tecnologías digitales.

La capacidad de los países del Sur global para participar en la formulación de normas internacionales sobre IA determinará en gran medida su influencia en el futuro del sistema digital global. La falta de representación en los foros internacionales de gobernanza digital podría resultar en la imposición de normas que favorezcan los intereses de las grandes potencias tecnológicas, dejando a los países del Sur global en una posición subalterna. Por ello, es importante que los países en desarrollo trabajen en conjunto para fortalecer su voz en estos debates, participando activamente en la creación de marcos regulatorios y éticos para el uso de la IA.

IA y hegemonía en disputa: China, los BRICS y el Sur global

Asimismo, la cooperación multilateral entre los BRICS es clave para garantizar que la IA se utilice de manera inclusiva y beneficiosa para todos los países miembros. Esta cooperación permite no solo el intercambio de conocimientos y tecnologías, sino también la creación de una red de solidaridad que desafíe la dominación de las grandes potencias tecnológicas del Norte. A través de esta red, los BRICS pueden compartir buenas prácticas, promover la investigación conjunta y establecer políticas comunes que favorezcan el desarrollo digital en sus respectivas regiones (Kovacevic, 2020). A medida que los BRICS se consolidan como un bloque tecnológico emergente, tienen la oportunidad de desempeñar un papel clave en la creación de un sistema digital global más equilibrado y justo.

El papel de China y la IA en la reconfiguración del poder tecnológico global presenta tanto oportunidades como desafíos para el Sur global. Por un lado, la colaboración con China y el acceso a tecnologías avanzadas pueden ayudar a los países en desarrollo a integrarse en la economía digital global y a reducir la brecha tecnológica. Sin embargo, este avance debe ir acompañado de políticas nacionales y regionales que fomenten la innovación local, fortalezcan las capacidades de investigación y desarrollo, y promuevan la creación de un ecosistema digital inclusivo. Además, los países del Sur global deben asegurarse de que sus voces sean escuchadas en los debates internacionales sobre la gobernanza de la IA, con el fin de garantizar que las normas globales favorezcan sus intereses y promuevan el desarrollo sostenible.

La inteligencia artificial, como campo de poder, exige ser pensada más allá de su dimensión técnica. Tal como advertía Maquiavelo con su metáfora del centauro, el ejercicio del poder requiere combinar la razón con la fuerza y la astucia. En ese sentido, los países del Sur global no sólo deben aspirar a dominar ciertas tecnologías, sino también a comprender las lógicas políticas y estratégicas que las sustentan. El desafío para el Sur global en la era de la IA es grande, pero también lo son las oportunidades. A medida que China lidera la carrera hacia la supremacía digital, los países en desarrollo tienen la posibilidad de asumir un papel activo en la creación de un futuro digital más equitativo, basado en la cooperación sur-sur y la innovación local. La clave estará en encontrar un equilibrio entre aprovechar las oportunidades que ofrece la IA y evitar los riesgos de dependencia tecnológica, al mismo tiempo que se construyen nuevas estructuras de gobernanza global que favorezcan una distribución equitativa del poder y los beneficios de la tecnología digital.

Referencias

Arrighi, Giovanni (2007). Adam Smith en Pekín: El ascenso de China y el futuro del capitalismo mundial. Akal.

- Birhane, Abeba** (2021). „Algorithmic injustice: A relational ethics approach“. *Patterns*, 2(2), 100205. Disponible en: <https://doi.org/10.1016/j.patter.2021.100205>
- Calderón, Alejandro y Harvey, David** (2023). „La reconfiguración de la hegemonía tecnológica global: El caso de China y los BRICS“. *Geopolitics Review*, 21(3), 142–160.
- Chakravartty, Paula y Zhao, Yuezhi** (Eds.). (2008). *Global communications: Toward a transcultural political economy*. Rowman y Littlefield.
- Couldry, Nick y Mejias, Ulises** (2019). *El costo de la conexión: cómo los datos colonizan la vida humana y la apropian para el capitalismo*. Stanford University Press.
- Descifrando la Guerra** (2024). „La Base 5x81 | DeepSeek: China contraataca en la carrera tecnológica de IA“. Disponible en: <https://www.youtube.com/watch?v=XXXXX>
- Ding, Jeffrey** (2018). *Deciphering China's AI dream*. Future of Humanity Institute, University of Oxford. <https://www.fhi.ox.ac.uk/china-ai-dream/>
- Dussel, Enrique** (2015). *Filosofías del sur: Descolonización y transmodernidad*. Siglo XXI Editores.
- Feldstein, Steven** (2019). „The global expansion of AI surveillance“. *Carnegie Endowment for International Peace*. Disponible en: <https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847>
- Financial Times** (2024). „Líderes tecnológicos de EE. UU. alarmados por el rápido avance de la IA en China“. Disponible en: <https://www.ft.com/>
- Gramsci, Antonio** (1971). *Selections from the Prison Notebooks* (Q. Hoare y G. Nowell Smith, Eds.). International Publishers.
- Harvey, David** (2003). *El nuevo imperialismo*. Oxford University Press.
- Harvey, David** (2005). *Espacios del capitalismo global: hacia una teoría del desarrollo geográfico desigual*. Franz Steiner Verlag.
- Harvey, David** (2014). *Diecisiete contradicciones y el fin del capitalismo*. Oxford University Press.
- Kovacevic, Jovan** (2020). „The BRICS and the digital economy: Challenges and opportunities“. *BRICS Journal of Innovation*, 3(2), 102–118.
- Lee, Kai-Fu** (2018). *Superpotencias de la IA: China, Silicon Valley y el nuevo orden mundial*. Houghton Mifflin Harcourt.
- Maquiavelo, Nicolás** (2008). *El Príncipe*. BN Publishing. (Edición en español). Capítulos XVIII–XIX contienen la referencia al centauro como metáfora del poder.
- Milan, Stefania y Treré, Emiliano** (2019). „Big data desde el Sur: más allá del universalismo de datos“. *Televisión y New Media*, 20(4), 319–335.
- Nye, Joseph** (2004). *Soft power: The means to success in world politics*. PublicAffairs.
- Ricaurte, Paola** (2019). „Epistemologías de datos, la colonialidad del poder y la resistencia“. *Televisión y New Media*, 20(4), 350–365.
- Santos, Boaventura de Sousa** (2010). *Epistemologías del Sur*. CLACSO.

- Saxena, Anika** (2023). „AI for all: Transforming India’s technological landscape“. *Indian Journal of Technology and Innovation*, 5(2), 45–67.
- Stuenkel, Oliver** (2016). *O mundo pós-ocidental. Potências emergentes e a nova ordem global*. Polity Press.
- Wallerstein, Immanuel** (2011). *El moderno sistema mundial I: La agricultura capitalista y los orígenes de la economía-mundo europea en el siglo XVI*. 2ª ed. aum. México: Siglo XXI Editores. Traducción de Antonio Resines. ISBN: 978-607-03-0337-1.
- Wallerstein, Immanuel** (2011). **El moderno sistema mundial II: El mercantilismo y la consolidación de la economía-mundo europea, 1600-1750**. 2ª ed. aum. México: Siglo XXI Editores. Traducción de Pilar López Máñez. ISBN: 978-607-03-0345-6.
- Wallerstein, Immanuel** (2011). **El moderno sistema mundial III: La segunda era de gran expansión de la economía-mundo capitalista, 1730-1850**. 2ª ed. aum. México: Siglo XXI Editores. Traducción de Jesús Albores. ISBN: 978-607-03-0354-8.
- Wallerstein, Immanuel** (2004). *Mundos y sistemas históricos*. Siglo XXI Editores.
- Zuboff, Shoshana** (2019). *La era del capitalismo de vigilancia: la lucha por un futuro humano en la nueva frontera del poder*. PublicAffairs.

Valoración crítica de la interacción Humano-IA en la construcción de sentido: Ejercicio en el marco de la Semiótica de las Pasiones.

Critical assessment of human-AI interaction in the construction of meaning: An exercise within the framework of the semiotics of passions.

Vanessa Márquez Vargas
vamv@ula.ve

Resumen

Este documento se aproxima a una valoración crítica de la interacción Humano-„Inteligencia Artificial“ (IA) desde la perspectiva de la Semiótica de las Pasiones. Se analiza cómo los grandes modelos de lenguaje reproducen patrones discursivos de un área específica de conocimiento, planteando un ejercicio comparativo entre respuestas de cuatro modelos de IA, lo que pone en evidencia la diferencia entre simulación estadística e interpretación significativa. La reflexión se sitúa en lo que se puede llamar un borde semiótico, donde la IA no sustituye el pensamiento humano, sino que confronta sus límites. Se insiste en la necesidad del pensamiento crítico para diferenciar entre simulación y comprensión, subrayando la importancia de mantener la dimensión pasional, experiencial y semiótico-hermenéutica del sujeto en el proceso de significación. El análisis concluye que la IA, aunque poderosa, no puede sustituir la sensibilidad encarnada ni los procesos semióticos vivenciales propios del humano, aun cuando proponga atractivos y creativos procesos de semiosis.

Palabras claves: semiótica de las pasiones, inteligencia artificial, interacción humano-máquina, significación, pensamiento crítico.

Abstract

This paper offers a critical assessment of the interaction between humans and „Artificial Intelligence“ (AI) from the perspective of the Semiotics of Passions. It analyzes how large language models reproduce discursive patterns from a specific area of knowledge, proposing a comparative exercise involving the responses of four AI models, which reveals the difference between statistical

simulation and meaningful interpretation. The reflection is situated at what may be called a semiotic boundary, where AI does not replace human thought but rather confronts its limits. The paper emphasizes the need for critical thinking to distinguish between simulation and comprehension, highlighting the importance of preserving the passionate, experiential, and semiotic-hermeneutic dimensions of the subject in the meaning-making process. The analysis concludes that AI, although powerful, cannot substitute embodied sensitivity or the lived semiotic processes unique to human beings, even if it proposes attractive and creative semiotic operations.

Keywords: semiotics of passions, artificial intelligence, Human-machine interaction, Meaning-making, Critical thinking.

1. A modo de introducción

El complejo entramado de la experiencia de lo humano motiva a que los distintos campos disciplinarios de las ciencias mantengan una constante tensión entre dos orientaciones epistemológicas fundamentales: por un lado, el valor de la tradición como sustrato vital de la condición humana —en la que están declaradas la socialización, la política, el arte, la cultura y todo mecanismo de creencia sustentado en la fe y, por otro lado, las posibilidades de „dominio“ científico sobre la totalidad del mundo conocido, dada la predicción sobre el universo en (re)conocimiento con un afán de progreso, cuyos límites se amplían en la medida en que se incrementan los estados de incertidumbre en torno a la realidad aparente de las cosas, la razón filosófica y la propensión lógica de la evolución del pensamiento y de los sistemas científicos y sociales en que se sostiene ese pensamiento.

En el campo de las ciencias sociales y ciencias humanas, la comprensión genuina o legitimadora de lo que compete al sujeto, a la persona, históricamente oscila entre dos polos; las reflexiones basadas en la utopía de la justicia social que cuestionan lo interno del sistema, entre espejismos que detonan revoluciones sociales y científicas, y las aproximaciones que, bajo el manto de „verdades“ parciales, impulsan al predominio de los principios del capital. Estas oscilaciones, a su vez, revelan los anhelos de verificabilidad de todo el conocimiento posible que tienen los investigadores de las ciencias en general, el esfuerzo cognitivo y las apuestas por la socialización en la producción de conocimiento, a partir del cuestionamiento y la respuesta a este cuestionamiento en la medida en que no hay un saber absoluto y que las formas de lo humano están configuradas tanto por la pregunta por el sujeto/objeto, como por aquello que lo rodea y se desprende él.

Quiere decir esto que las consideraciones científicas y sociales apuntan hacia varias direcciones y encrucijadas; en las que, entre otros caminos, se sigue hacia el dominio cada vez más técnico, tecnológico de los espacios comunes en la sociedad; hacia la materialización de herramientas para uso y consumo, envueltas por el manto de la manipulación y la prepotencia humana, supeditada al cálculo y a la racionalidad; lo cual recrea, como otros tantos eventos humanos, que la prevalencia de la arbitrariedad del deseo en la cadena evolutiva responde a impulsos sociales paradójicos. Por ejemplo, es paradójica la tensión entre el ideal de conocimiento de acceso libre y las condiciones materiales actuales que requieren millonarias inversiones de capital para el desarrollo tecnológico, patrocinado fundamentalmente por corporaciones privadas ubicadas en lo que en la nueva geopolítica se conoce como Norte Global en oposición, desde luego, al Sur Global en el que muchos países buscan consolidar un orden „diferente“ en las relaciones sociopolíticas.

En este sentido, el campo de las ciencias informáticas, en la que se trabaja desenfadadamente por crear y alimentar sistemas capaces de resolver y aliviar tareas humanas a gran escala, resulta atractivo y altamente polémico desde la perspectiva crítica de las ciencias sociales y ciencias humanas al plantear desafíos epistemológicos y éticos que, de cierto modo, revelan tensiones conceptuales, como las resistencias al cambio, los estados de conciencia, dualismos o idealismos filosóficos que no son letra muerta, pero sí son territorios de pensamiento que constituyen espacios de reflexión crítica sobre las implicaciones sociales de estos desarrollos, aunque no determinen directamente las decisiones en el ámbito tecnológico.

El hecho de que grandes modelos de lenguaje (LLM por sus siglas en inglés) sean capaces de producir, en segundos, textos coherentes desde el punto de vista gramatical, analizar contenido, reproducir, imitar estilos creativos e incluso presentar juicios estéticos o decisiones sobre actuaciones y conductas humanas, entre muchas otras cosas, constituye una ruptura significativa con las expectativas tradicionales sobre las capacidades exclusivamente humanas en estos dominios, lo cual genera desencuentros y temores, sobre todo en espacios académicos humanistas.

Dado que los modelos estadísticos procesan grandísimas cantidades de datos para „consumar“ nuestros deseos, cual genio saliendo de la lámpara, se producen estados de crisis en las formas conocidas del sistema social y, al mismo tiempo, nuevas paradojas que van y vienen como las olas del mar, evidenciando cómo se reconfiguran las nociones de identidad y se abordan parcelas de poder en las que oponentes y seguidores del avance del conjunto llamado „Inteligencia Artificial“ ven promesas y amenazas indistintamente.

Esta situación establece una dinámica compleja entre estados de producción técnica e intelectual que tienen muchas implicaciones debatibles en distintos

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

campos, más allá del contexto „aterrador“ para los humanos de ser sustituidos y dominados por las propias máquinas „inteligentes“ que han creado; están los vacíos históricos, las luchas incansables de las minorías por ser visibilizadas y ficciones que ahora mismo superan la dimensión de la „realidad“ que conocemos. Con esto, como señala Matteo Pasquinelli (2022, 17) refiriéndose a la crítica hecha por Cathy O'Neil, „está quedando claro que el problema de la IA no tiene nada que ver con la inteligencia en sí misma, sino con la forma en que se aplica a la gobernanza de la sociedad y al trabajo a través de modelos estadísticos que deberían ser transparentes y expuestos al escrutinio público“.

El boom de la „Inteligencia Artificial“ que se vive en la actualidad sitúa al sujeto social y particularmente al sujeto intelectual, académico, investigador, frente a un espejo que proyecta múltiples versiones de una misma imagen, todas atractivas e igualmente cuestionables en el marco de las aspiraciones, ansiedades y contradicciones sociales que operan, por decirlo de algún modo, sobre el escenario de combinaciones y recombinaciones algorítmicas y estadísticas de patrones preexistentes de conocimiento y carencia en el mismo nivel, debido a que el control sigue estando en manos del titiritero y la „Inteligencia Artificial“, pese a que puede tener muchos aspectos positivos que destaca muy bien la burbuja de marketing en la que viaja, es también reflejo de la desigualdad en las políticas globales y conflictos sociales (Alex Avila, 2025). Pues, aunque pueda parecer una obviedad, los modelos de lenguaje, la „Inteligencia Artificial Generativa“ que se está presentando a los ojos del mundo no surge de un vacío tecnológico en un nicho de conocimiento, sino como respuesta a necesidades situadas en transformaciones socioculturales históricas y en una idea de supremacía del hombre por el hombre que ahora se torna hombre-máquina.

En el ámbito de las ciencias sociales y humanas, la „Inteligencia Artificial“ se puede ver como un asistente infalible y prometedor que ilumina el camino hacia las profundidades de la lógica inductiva, deductiva y el discernimiento (Pizarro-Romero, J., y Lovón, M, 2025). Sin embargo, es importante llamar la atención e insistir en ello; en que la „Inteligencia Artificial“ es estadística computacional, la arquitectura de una red neuronal entrenada para procesar datos a gran escala; un gran modelo de lenguaje, conjuntos de sistemas basados en aprendizaje humano, lo cual permite que la máquina sea extremadamente eficiente y rápida en las tareas que le soliciten; pero el conocimiento humanístico requiere interpretación contextual, juicio valorativo y comprensión experiencial que no pueden reducirse solo a patrones algorítmicos y, al menos por ahora, requiere del humano.

Por tal motivo, una de las tantas razones por las cuales el uso de „Inteligencia Artificial“ resulta „alarmante“ en ciencias sociales y en ciencias humanas, en ámbitos específicos como el educativo o artístico, por ejemplo, es porque se asumen a los sistemas inteligentes como interlocutores deslumbrantes al simular, interpretar, generar y „negociar“ significados en diversos campos del

conocimiento humano que al sujeto, a la persona común, le tomaría o le ha tomado años adquirir y comprender; años de lectura, reflexión, razonamiento, contrastación de experiencias que no son suficientes para que un solo sujeto abarque el todo. Por esta razón es abrumadora la experiencia de interacción que se da entre el humano y los sistemas inteligentes, debido a la capacidad de estos sistemas-máquinas para almacenar y procesar datos en los más variados escenarios de la vida social y de las disciplinas científicas, artísticas, políticas —pensar en áreas como la medicina o el derecho implicaría otros caminos reflexivos que no son el caso de estudio en este momento—.

En vista de esto, es importante prestar atención a la interacción humana con „Inteligencia Artificial“ no como un simple intercambio informacional, sino como un proceso semiótico complejo donde se proponen distintos niveles de agenciamiento de sentido. Reconociendo que la „Inteligencia Artificial“ realiza operaciones semióticas sofisticadas —mediante la codificación de contextos y la conexión entre símbolos—, estas operaciones difieren cualitativamente de la semiosis experiencial humana caracterizada por la copresencia sensible. La „Inteligencia Artificial“ puede procesar y correlacionar significados, reproduciendo incluso patrones de sensibilidad —si se comprende la semiosis como lo enunció Peirce en principio y quienes han ampliado el horizonte de sus ideas como Eco, Verón, Lucifora, entre otros—, aunque la reproducción de sensibilidad no equivale a la experiencia sensible que siente el sujeto en su vida.

Esto más allá de invalidar las capacidades semióticas de la „Inteligencia Artificial“, permite comprender modalidades específicas de operación de los sistemas; porque más allá de ser consideradas como semiosis „artificiales“ las operaciones de la máquina aportan al propio sujeto una experiencia de intercambio valiosa. Como destaca Maria Clara Lucifora (2021, 28):

„[...] desde las humanidades, debemos profundizar este tipo de análisis y promover el desarrollo del pensamiento crítico para que la consideración de los modelos y los softwares de IA sea minuciosa tanto por parte de sus creadores como de sus usuarios, con el fin de reducir al mínimo cualquier tipo de arbitrariedad que devenga en injusticia y sufrimiento [...]“

En este marco, la valoración de la interacción entre los humanos y la „Inteligencia Artificial“ puede hacerse siguiendo modelos fundamentados en teorías de las ciencias del significado, con las cuales se puede observar cómo se teje la red en la que se sostienen los cambios desde el punto de vista lingüístico y, por tanto, desde el punto de vista histórico, cultural y social. Desde esta perspectiva, también es posible acercarse a las propuestas de Eric Landowski (2004, 2017) para considerar la interacción humana con sistemas de „Inteligencia Artificial“ como modalidades específicas de interacción significativa que combinan elementos de programación y manipulación, en las cuales la máquina reproduce patrones complejos de significación sin

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

experimentar la dimensión vivencial que caracteriza. Por ejemplo, la responsabilidad semiótica y el sentido atribuido a lo dicho en un marco específico de acciones, particularmente de emociones o pasiones.

Esto no implica una limitación absoluta, sino una diferencia operacional. La máquina es ese „otro“ que desencadena en los usuarios del sistema, estados particulares que van del éxtasis ante la aparente creatividad y coherencia de las respuestas generadas, hasta la frustración por outputs no verificables o llamadas „alucinaciones“ del sistema, que constituyen salidas originales sin conexión directa con los datos de entrenamiento y que pueden ser útiles como actos creativos, independientemente de las expectativas del usuario. En un lenguaje más técnico la explicación sería, siguiendo a Aandkumar citada por Greengard (2025):

„A medida que las palabras se convierten en vectores numéricos y los valores numéricos se multiplican por un orden de magnitud dentro de un modelo, las cosas se vuelven confusas y los significados se difuminan. El modelo simplemente regurgita palabras basándose en la probabilidad. „Las alucinaciones de la IA son una característica, no un error [...] Los modelos se ven incentivados a producir texto plausible en lugar de texto factual“¹.

Los sistemas de „Inteligencia Artificial“ tienen una capacidad de memoria superior a la de cualquier humano y esta es una de las características que hace de estos sistemas en extremo relevantes para los desarrollos tecnológicos y al mismo tiempo es punto de quiebre en el que la crítica hace énfasis y sobre el cual se trabaja para superar límites.

Si bien es cierto que, entre otras cosas, los sistemas identifican y recuperan fragmentos textuales asociados con autores, conceptos y teorías relevantes, proporcionan síntesis conceptuales destacadas y adoptan una posición enunciativa que simula competencia disciplinar y distancia crítica a petición del usuario; también es posible que generen expansiones analógicas motivando indistintamente insights creativos y distorsiones conceptuales significativas que resultan adversas para los aprendizajes de usuarios aparentemente ingenuos que confían ciegamente en que el sistema arroja „verdades“ reflexivas, cuando ante lagunas de conocimiento específico revisado y calificado, las máquinas tienden a procesar información de distintas perspectivas teóricas como si fuesen complementarias o fácilmente integrables, ignorando las tensiones epistemológicas fundamentales que

¹ **En si idioma original:** „As words become numeric vectors and numerical values multiply by an order of magnitude inside a model,i things get fuzzy and meanings get blurred. The model simply regurgitates words based on probability. “AI hallucinations are a feature rather than a bug... Models are incentivized to produce plausible text rather than factual text“.

existen entre distintas tradiciones teóricas, escuelas, corrientes y movimientos científicos-estéticos.

Esto sucede, al menos hasta el momento, con versiones de prueba o versiones gratuitas de sistemas de „Inteligencia Artificial“, porque el sistema no puede realmente „reconocer“ el sentido de sus propias producciones ni comprender plenamente las condiciones sociales e institucionales que determinan su recepción. Los procesos de entrenamiento del sistema son complejos y se nutren de enormes repositorios de datos que constituyen corpus generales, aunque la personalización contextual ocurre a nivel de consulta mediante prompts específicos.

Esta dualidad entre generalidad del modelo y especificidad contextual revela que hay múltiples actores involucrados en los resultados que obtenemos de una „Inteligencia Artificial“ después de una consulta; desde curadores de datos, ingenieros especializados, hasta el usuario final, incluyendo los sistemas de retroalimentación que continuamente modifican las respuestas futuras y conforman el sistema de predicción.

Al respecto, es importante considerar que los sistemas de „Inteligencia Artificial“, si bien aprovechan el orden secuencial de palabras en sus textos de entrenamiento, al combinar y recombinar estadísticamente múltiples datos son capaces de generar espacios semióticos que trascienden esta secuencialidad original para crear mundos semióticos a partir de nuevas yuxtaposiciones de signos, desarrollando sistemas de referencia que, aunque se nutren de patrones existentes, generan configuraciones que nunca existieron históricamente; procurando, al mismo tiempo, variaciones dentro de un sistema de expectativas como señala Eco (1992). Es decir, una lógica interna de estos mundos que refleja distribuciones probabilísticas en lugar de intencionalidad en tanto concepto formalmente conocido.

Cabe señalar que la noción de intencionalidad ha generado profundos debates en el ámbito de la filosofía del lenguaje y la lingüística. Sobre este constructo, entre otros, Putnam (1990) sostuvo que el significado trasciende la voluntad individual del hablante porque la intencionalidad es un fenómeno emergente que surge de la interacción compleja entre factores cognitivos, textuales y sociales, refiriéndose a las interacciones humanas en las cuales los hablantes participan de la referencia del objeto, sin dominar necesariamente los criterios de aplicación conceptual del campo científico que define a ese objeto.

Esto es fundamental, debido a que en los últimos días las disputas conceptuales al respecto de la intencionalidad plantean la cuestión de si estas distribuciones probabilísticas de la „Inteligencia Artificial“ podrían arrojar formas emergentes de intencionalidad y los ejemplos saltan a la luz pública

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

con las declaraciones de los propios desarrolladores. En un reciente reportaje McMahon (2025) señaló que durante las pruebas de actualización y lanzamiento de Claude Opus 4 es visible cómo el sistema es capaz de defenderse chantajeando a los ingenieros frente a una posible desactivación del sistema. Con lo cual se hace necesaria una reconceptualización y reconocimiento de continuidades y discontinuidades teóricas que se pueden hacer en la medida en que el usuario esté capacitado para notar las sutilezas de cambio en las proposiciones que arroja el sistema ante una consulta.

Los sistemas están en aprendizaje continuo y nos están diciendo que no sólo procesan corpus estadísticos estáticos, sino que evolucionan a través de las interacciones con los usuarios. Esto sugiere que la relación humano-sistema es constitutiva y cada interacción modifica potencialmente las futuras respuestas del sistema. Al mismo tiempo, esto es significativo como un punto álgido ante la mirada crítica, en tanto, es impactante para el usuario la capacidad del sistema para construir y simular representaciones de experiencias que nunca sucedieron, al mejor estilo de los mundos posibles ficticiales siguiendo principios de verosimilitud.

Asimismo, la capacidad de generar simulaciones de estados afectivos mediante correlatos lingüísticos sin estar anclados en experiencias corporales vividas y sentidas. De modo que se dificulta, en cierto modo, mantener la disciplina interpretativa que procura la mirada objetiva general de las ciencias, ya que cada vez es más difícil superar la tendencia humana a antropomorfizar los objetos y muy especialmente a las máquinas inteligentes.

Desde el punto de vista de las interpretaciones fenomenológicas o las interpretaciones tendientes a la subjetividad, también se plantean retos para explicar al usuario, fundamentalmente, que las respuestas o salidas a preguntas puntuales, o solicitudes demandantes en áreas específicas de conocimiento, son el resultado de procesos estadísticos complejos y no de estados intencionales cooperativos naturales de los procesos de interacción y socialización entre humanos.

Pese a la *verdad de Perogrullo* que esto puede suponer, la utilidad y veracidad científica de las salidas o respuestas que arroja un sistema de „Inteligencia Artificial“ están fuertemente condicionadas por el conocimiento previo del usuario y su capacidad crítica, en tanto las respuestas pueden crear una ilusión de comprensión que oculta la complejidad real de los conceptos y proposiciones teóricas abordadas.

Por ejemplo, la aparente exhaustividad bibliográfica sobre un tema o área de conocimiento alimenta una erudición y una profundidad conceptual que resulta una simulación en tanto el sistema se adapta a las expectativas del usuario, reforzando con ello sesgos interpretativos en vez de propiciar cuestionamientos epistemológicos significativos sobre la razón social de los

productos científicos, por decir lo menos; a lo que se puede llamar en palabras de Verón (1996) un „efecto de reconocimiento“ o de naturalización de interpretaciones motivadas por posicionamientos culturales e ideológicos en el marco del entramado discursivo de la construcción social del sentido.

El potencial „explicativo“ de los sistemas de „Inteligencia Artificial“ seduce a primera vista por lo avasallante de los datos recuperados a gran velocidad, como se ha mencionado. Sin embargo, el procesamiento algorítmico puede verse disminuido —hasta tanto se le den nuevos suministros de datos— ante la dimensión fenomenológica de la interpretación que, para el momento, sólo es posible entre humanos; pues aunque el sistema puede reproducir comportamientos que sugieren competencia comunicativa, a partir de secuencias de patrones estadísticos extraídos de corpus textuales, incluso manteniendo coherencia temática durante interacciones prolongadas; se pierde en un punto la dimensión normativa y autoconsciente de la competencia semiótica humana.

Es decir, la salida del sistema arroja conceptos elaborados que se interceptan generando versiones explicativas sobre los aspectos solicitados pero de modo superficial, para compensar al usuario pero sin transformar el proceso en una interpretación/comprensión en sentido semiótico-hermenéutico, independientemente de la duración o complejidad aparente de la interacción.

Todo esto, en suma, muestra que la comprensión de las fases que anteceden y siguen a la significación son relevantes para reflexionar críticamente sobre la actuación e interacción humano-sistemas „Inteligencia Artificial“ en tanto que desde el punto de vista semiótico, la significación es un proceso activo de construcción en una dimensión holística del significado y no sólo de decodificación de signos.

Asimismo, es relevante porque permite ver dos cosas; la primera de ellas es que los sistemas de „Inteligencia Artificial“ son capaces de fundar nuevos caminos epistemológicos basados en coherencia estadística, aunque estos caminos están condicionados desde el inicio por los sesgos del sustrato que alimenta los datos, limitando con ello la potencialidad interpretativa desde perspectivas horizontales a favor de los sistemas sociales ampliados, colectivos y colaborativos.

La segunda cosa es que se pueden aprovechar las fisuras y las inconsistencias de los sistemas inteligentes para analizar qué aspectos del conocimiento humanístico se hacen visibles y cuáles se invisibilizan, atendiendo especialmente a las asimetrías de poder que subyacen a la producción y control de la tecnología como tema político y sociológico fundamental que ayudaría a superar „el miedo“ a seguir viviendo los horrores de un mundo desigual y polarizado, más que el miedo a la máquina,

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

reconociendo que incluso la pregunta sobre si estos sistemas poseen propósitos genuinos permanece epistemológicamente compleja dada la opacidad de su funcionamiento interno.

Al respecto son notables las proposiciones críticas de Pasquinelli (2019, 2022, 2023) en su detallada explicación sobre modelos inteligentes, en las que enfatiza la vulnerabilidad, sesgos y falacias que pueden tener consecuencias sociales de alto impacto; lo más interesante, es que al señalar errores o inconsistencias de los sistemas inteligentes, pérdidas de información y problemas de interpretación, Pasquinelli también resalta la intencionalidad del engaño ante decisiones políticas y sociales erradas en un estado de anomalía social persistente. Esto es notable cuando los grandes representantes del capital hacen una apuesta casi total en términos económicos por el desarrollo de modelos eficientes de aprendizaje automático y a su vez los científicos declaran preocupación por los detalles que marcan diferencias en la evolución de la máquina.

En entrevista reciente Block (2025) habló sobre los errores y los motivos por lo que ChatGPT se equivoca ante ejercicios puntuales de prueba cognitiva, „lo más llamativo es que estos errores no son anecdóticos, sino estructurales“, en este mismo sentido, salvando las distancias, se puede pensar que los errores aparentemente más simples de la sociedad en pleno siguen siendo los más difíciles de corregir y la burbuja comercial de la „Inteligencia Artificial“ está alimentando a un gran monstruo; al respecto Pasquinelli (2019, 24) apunta:

„[...] la actual carrera armamentística entre las empresas de IA sigue consistiendo en encontrar los algoritmos más rápidos para calcular modelos estadísticos. La comprensión de la información, por tanto, mide la proporción de ganancias de estas empresas, pero también, la proporción de pérdida de información – y dicha pérdida suele significar una pérdida de la diversidad cultural del mundo [...]“

Esto no es nada nuevo si pensamos en la historia contemporánea haciendo repaso sobre los llamados de atención y discusiones de estudiosos de diversas disciplinas, no solamente sobre aspectos científicos-tecnológicos sino políticos, tecnocráticos, geopolíticos y sociales acerca de la polarización, el tratamiento desigual, las revoluciones o los genocidios, lo que se resume en el manejo del poder y el control social. Sobre estos temas hay una vasta cantidad de bibliografía vinculante de acuerdo a la perspectiva que se quiera seguir y sobresalen los nombres de Gilles Deleuze, Michel Foucault, Noam Chomsky, Jacques Derrida, Ned Block, apuntando hacia la idea de desnaturalizar la tecnología y situar el desarrollo de modelos inteligentes como parte de la reconfiguración social.

Cada uno de estos estudiosos en su momento problematizó y problematiza situaciones que se mantienen vigentes. El ejercicio del poder con la operación

de dispositivos de control que se asemejan a la inteligencia algorítmica para modelar y moldear comportamientos y deseos traduciéndolos a necesidades; manufactura de consentimiento no informado para obtener datos sensibles de la persona a través del consumo mediático, la generalización entre la estructura y la técnica con relación a la producción textual y, entre muchas cosas más, las críticas y dudas sobre la posibilidad de que a futuro los sistemas de „Inteligencia Artificial“ posean verdadera comprensión o experiencia subjetiva (Block, 2025).

Sin embargo, más allá de los presagios, lo dicho por Pasquinelli (2019, 2022) se puede asumir también como una invitación a mirar con cuidado el proceso evolutivo de los sistemas de „Inteligencia Artificial“, no como un fin en sí mismos, sino como una herramienta cuyo valor debe estar volcado al servicio de la inteligencia humana como soporte, como puente entre diferentes sistemas simbólicos, pues lejos de ser infalible la „Inteligencia Artificial“ funciona tanto para potenciar el estudio de sistemas semióticos debido a la capacidad de la máquina para procesar múltiples lenguajes y códigos, como para identificar sus limitaciones estructurales sobre temas y áreas específicas en las que se requiere tanto conocimiento humano especializado, como reconocimiento sensitivo del entramado de las experiencias humanas en la dimensión pasional y afectiva; tal es el caso de las Semiótica de las Pasiones en el marco de la Semiótica como constructo teórico y científico.

Para mostrar con más claridad estas tensiones y distensiones se presenta un sencillo ejercicio comparativo que consiste en consultar a 4 modelos diferentes de „Inteligencia Artificial“ una misma pregunta que se considera compleja en el núcleo metodológico de la Semiótica de las Pasiones, con base en un texto en forma de relato con pretensión literaria; con el objetivo de analizar las variaciones, semejanzas y puntos ciegos entre las respuestas generadas y las formas de presentación de respuestas. Esto permite observar y reflexionar acerca de cómo cada sistema interpreta y procesa la información, para notar diferencias o similitudes en los resultados o salidas de las bases de datos y algoritmos de entrenamiento, así como también las limitaciones inherentes a la comprensión de la máquina sobre el contexto cultural y semiótico específico que se pide.

A diferencia a las proposiciones estructurales de lo que se conoce como Semiótica de Primera Generación, la Semiótica de las Pasiones, o de Segunda Generación, Semiótica de la acción, constituye un campo de conocimiento que involucra la dimensión pasional del sujeto, relaciones experienciales, corporales y afectivas que requieren comprensión de los procesos de significación sensitiva y su manifestación discursiva. En vista de esto, no solamente es necesario el conocimiento teórico estructural para abordar un texto y configurarlo en sus dimensiones lingüísticas formales, sino que también es necesaria la comprensión de los matices; es decir, el cómo las

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

pasiones se actualizan en configuraciones discursivas específicas para captar las modalidades pasionales, las gradaciones tensivas y las transformaciones del sujeto pasional, narrativo y enunciante.

En este sentido, con el ejercicio se evidencia hasta qué punto, hasta el momento, los modelos consultados pueden o no captar la sutileza que reclama la Semiótica las Pasiones. Porque una cosa es procesar estructuras narrativas básicas o identificar actantes en un esquema funcional, y otra muy distinta es percibir los elementos tensivos que atraviesan un discurso (Fabbri, 2000). La intención es mirar las respuestas que generan los diferentes modelos, las estrategias interpretativas y la presentación de la información pensando en los límites de la dimensión fenomenológica que exige una sensibilidad semiótica y hermenéutica humana.

En el ejercicio se plantea la consulta a ChatGPT-4, Gemini 2.0, Claude Sonnet 4, y DeepSeek, utilizando como prompt inicial,

2. ¿Cómo operan los estados sensibles en la configuración de la narración?

Una pregunta relevante en la teoría de la Semiótica de las Pasiones de Greimas y Fontanille (1994), con pequeñas derivaciones para mantener la consistencia de las respuestas en el procesamiento de la información teórica básica (Greimas y Julius, 1971 y 1973). A partir de las respuestas obtenidas, se les solicitó a los modelos que volcaran ese conocimiento teórico desplegado hacia la valoración de un relato para ver la transición desde el manejo conceptual abstracto hacia la aplicación interpretativa concreta. Seguidamente, se solicitó realizar un análisis del texto por categorías, partiendo específicamente de las propuestas metodológicas de la Semiótica de las Pasiones, con la finalidad de identificar cómo cada „Inteligencia Artificial“ plantea las modalidades pasionales, los componentes tensivos y transformaciones del sujeto pasional dentro del entramado narrativo.

A continuación se presentan los resultados o salidas de modelos consultados ChatGPT-4, Gemini 2.0, Claude Sonnet 4, and DeepSeek respectivamente. Las tablas muestran la selección y organización en la que cada uno de los modelos configuró el análisis. Es importante destacar que al final de la respuesta en formato tabla, el modelo también presenta un comentario que denomina „Ajustes claves“, „Claves teóricas“, „Notas“. Esto permite, a ojos expertos, ver claramente el plano de inmanencia textual en el que operan estos sistemas simulando estructuras argumentativas que, en este caso particular, no son vinculantes, por lo tanto se pueden considerar erróneas. Pues, aunque las propuestas de filósofos como Deleuze y Rancière se corresponden con intereses cuestionadores de la teoría del arte o filosofía política, no tienen un perfil directo de actuación sobre la teoría de la Semiótica

de las Pasiones, de esta forma el análisis por categoría que los sistemas arrojan como resultado está falseado.

Los sistemas reproducen metalenguaje técnico, lo categorizan y, aunque sea errado, presentan un análisis y con ello satisfacen el deseo primario del usuario que es obtener una respuesta a la pregunta del prompt. Es en este punto donde la persona debe completar las lagunas de conocimiento y corregir las imprecisiones, con lo cual se hace imprescindible que no se confunda la simulación estadística del discurso con la comprensión conceptual y mucho menos con la interpretación. Puesto que la simulación estadística opera mediante reconocimiento de patrones y generación probabilística de respuestas y la comprensión conceptual requiere apropiación crítica del conocimiento y capacidad de establecer relaciones significativas entre conceptos. En este sentido, la interpretación añade la dimensión hermenéutica que involucra mediación subjetiva y experiencia vivida para que los procesos de asimilación superen la reproducción de estructuras discursivas y se vuelvan conocimiento.

La interpretación en tanto proceso humano está cargada de matices y complejidades. Y no depende únicamente del conocimiento teórico sobre un área; no se trata de resumir ideas principales de un texto para explicar terminología específica (García Contto, José David). Cuando el proceso interpretativo se genera sobre un texto, o en cualquier situación humana, el resultado está mediado por la subjetividad, la experiencia y la relación que se tenga con el mundo cultural; se generan perspectivas que pueden transformarse con el tiempo. La interpretación y la significación pasional surgen de la experiencia encarnada del sujeto. La dimensión pasional resulta inimaginable de producir por los sistemas que carecen de cuerpo y de experiencia fenomenológica del mundo —al menos hasta ahora, porque la evolución de estos sistemas avanza mucho más rápido que la capacidad humana de procesar el acontecimiento—.

Sin embargo, los sistemas inteligentes simulan estas experiencias pero con distancias que, aunque no siempre evidentes, hacen necesario el desarrollo de herramientas analíticas que permitan identificarlas, al ojo humano. Estas distancias permiten ver que la máquina combina un régimen de sentido con elementos de programación y manipulación (Landowski, 2004, 2017). La responsabilidad interpretativa recae en el sujeto, quien debe desarrollar la capacidad crítica para distinguir entre simulación y comprensión genuina, reconociendo que los sistemas operan bajo parámetros de diseño preestablecidos.

Los datos de entrenamiento en los cuales se sostiene la arquitectura de los sistemas de „Inteligencia Artificial“ están orientados claramente a persuadir al usuario. Si bien es cierto que se consultan estos sistemas sobre áreas que

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

desconocemos, en un problema de recuperación de conocimiento, esto no invalida que los datos de entrenamiento han sido configurados para maximizar la satisfacción del usuario a través de respuestas que parezcan completas, „reales“ incluso cuando no lo sean.

Un prompt contiene los „deseos“ del usuario y al mismo tiempo también puede sesgar las respuestas hacia direcciones específicas. Sin embargo, el sesgo determinante es el que ya porta la red neuronal debido a sus datos de entrenamiento, no el que introduce el prompt individual de cada usuario. El prompt opera como contexto que será procesado a través de sesgos preexistentes en la arquitectura del sistema (Theobald, 2024).

Esto se traduce en un llamado de atención hacia la satisfacción o frustración de posicionamientos éticos que históricamente la sociología y la filosofía han estudiado, especialmente las implicaciones políticas sobre lo que se enuncia en la sociedad como verdad, o realidad y la actuación del sujeto frente a la tecnología mediada por relaciones asimétricas de poder (Foucault, 1978). Estos sesgos estructurales operan de manera menos visible pero más determinante que los sesgos individuales del prompt, reforzando las relaciones asimétricas de poder.

De este modo ahora el sujeto, la persona, no solo está condicionada solamente por los factores políticos, sociales y culturales de su tiempo de socialización y en correspondencia con los desvaríos de la historia contemporánea, sino también por la utilidad y los avances tecnológicos y la postura que asuma ante la convivencia directa con sistemas de „Inteligencia Artificial“. Este condicionamiento se da porque el sujeto se encuentra situado en un entramado de mediaciones que configuran sus posibilidades de comprensión y acción. No se trata de determinismo, sino de reconocer que nuestras formas de acceder al conocimiento, de interpretar la realidad y de posicionarnos éticamente están ahora atravesadas también por la mediación tecnológica. Esta nueva capa de condicionamiento no elimina la voluntad del sujeto, pero sí añade una dimensión adicional de complejidad a los procesos de subjetivación en todo el orden de la vida social cotidiana.

Presentación resumida del Ejercicio

Los sistemas ChatGPT-4, Gemini 2.0, Claude Sonnet 4, y DeepSeek, generaron categorizaciones mediante el reconocimiento de patrones discursivos asociados al metalenguaje de la semiótica, combinando terminología técnica específica con estructuras tabulares que aparentan sistematicidad analítica. Sin embargo, las referencias a conceptos como „agenciamientos“, „régimenes de signos“, „cuerpo sin órganos“ de Deleuze, o „política de lo sensible“ de Rancière no tienen relación directa con la Semiótica de las Pasiones como marco teórico específico desarrollado por Greimas y Fontanille (1994).

Los sistemas armaron su propio estado de categorías sin precisar verdaderamente la teoría solicitada y los cuatro sistemas lo hicieron de modo similar. La Semiótica de las Pasiones opera desde conceptos específicos que configuran el significado en el paso de la virtualización a la realización de los estados pasionales. Y en vez de articular estos fundamentos teóricos precisos, los sistemas combinaron referencias filosóficas heterogéneas que aparentan ser académicamente convincentes por que no corresponden al aparato conceptual de la semiótica greimasiana.

En este sentido, la finalidad de estas tablas, del ejercicio propiamente, es servir como herramienta de análisis textual, organizando las categorías semióticas de manera que permitan comprender cómo los estados sensibles configuran la narración del cuento. Sin embargo, al examinar los resultados, es evidente que los sistemas de „Inteligencia Artificial“ produjeron esquemas categoriales reconocibles sin desarrollar una comprensión conceptual profunda de la Semiótica de las Pasiones como marco teórico específico.

Tabla 1

Análisis de Datos por Categorías: Semiótica de las Pasiones en la Configuración Narrativa

Pregunta de Investigación: ¿Cómo operan los estados sensibles en la configuración de la narración?

Categoría	Subcategoría	Descripción/ Operación	Evidencia del Cuento „El viaje de la brisa“
Construcción de Atmósfera	Tono emocional dominante	Estados sensibles definen el clima afectivo del relato (melancolía, euforia, tensión).	„La brisa se tornó fría, atrajo nubarrones y provocó una fuerte lluvia“ (tristeza transforma el entorno).
	Espacialidad sensible	Los espacios se cargan de emociones (lugares opresivos, liberadores, nostálgicos).	„La casa pequeña... con olor a clavos y canela“ (atmósfera cálida y hogareña).
Dinámica de Personajes	Emociones como motor narrativo	Decisiones y acciones guiadas por estados sensibles.	La brisa ayuda al bebé a respirar por compasión: „entró por su nariz... el niño lloró“.
	Epifanías sensibles	Momentos de revelación vinculados a percepciones afectivas.	„Comprendió que no es en el cuerpo donde están los sentimientos“ (epifanía en la muerte).

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

Empatía y Conexión	Sintonía afectiva con el lector	El texto evoca emociones compartidas (dolor, alegría, nostalgia).	„Jugó con unos niños que se reían“ (alegría contagiosa).
	Corporalidad narrativa	Sensaciones físicas que generan identificación.	„Refrescándoles el cuerpo [a los obreros]“ (tacto como acto empático).
Ritmo y Temporalidad	Progresión emocional	La trama avanza según pulsos afectivos (aceleración, pausa).	„Se quedó enredada en las hojas de palmeras hasta el atardecer“ (ritmo contemplativo).
	Tiempo subjetivo	Duración percibida según estados internos (dilatación, compresión).	„El hombre exhaló por última vez“ (momento fúnebre ralentizado).
Política de lo Sensible	Visibilización de lo invisible	Dar voz a experiencias marginales o silenciadas.	La brisa percibe el dolor del parto y la muerte: „gotas de sudor... el alma quería liberarse“.
	Resistencia afectiva	Emociones como actos de desafío o transformación.	„Arrojó tierra a los ojos del atacante“ (ira como justicia).

Claves Teóricas:

1. Semiótica de las pasiones (Deleuze): Los estados sensibles son „agenciamientos“ que conectan cuerpos, espacios y significados.
2. Política de lo sensible (Rancière): La brisa revela lo invisibilizado (parto, muerte) como acto político.
3. Ejemplo destacado: La escena de la anciana („como el director de una orquesta“) muestra cómo lo sensible construye comunidad.

Nota: Las categorías integran lo académico (teoría narrativa) y lo creativo (análisis textual). ¿Necesitas ajustar enfoques o añadir más ejemplos?

Nota: Categorías de Análisis (basadas en la semiótica de las pasiones y los documentos proporcionados) como salida al *prompt* ChatGPT-4

Tabla 2

Análisis de Datos por Categorías: Semiótica de las Pasiones en la Configuración Narrativa

Pregunta de Investigación: ¿Cómo operan los estados sensibles en la configuración de la narración?

Categoría	Subcategoría	Descripción/ Operación	Evidencia del Cuento „El viaje de la brisa“
Inmersión Sensorial	Apelación a los cinco sentidos	Uso de descripciones multisensoriales para crear realismo.	„El olor dulce a clavos, canela y jengibre impregnaba el lugar“ (olfato). „Aire helado que erizaba la piel“ (tacto).
	Construcción de atmósfera	Los sentidos vinculados a emociones específicas (nostalgia, tensión).	„Provocó una fuerte lluvia; la brisa se tornó fría“ (tacto + tristeza).
Emoción y Memoria	Evocación de estados afectivos	Detalles sensoriales como disparadores de respuestas emocionales.	„Recordó el olor del Atlántico, a mar salado“ (olfato + nostalgia).
	Simbolismo sensorial	Objetos o sensaciones que adquieren carga emocional.	„El pequeño papel con el número de teléfono“ (tacto + esperanza).
Mostrar vs. Contar	Lenguaje concreto y experiencial	Sustitución de abstracciones por percepciones sensibles.	„El sabor amargo que le quemaba la lengua“ vs. „estaba disgustado“.
	Acciones sensoriales como caracterización	Los personajes se definen por su interacción sensorial con el mundo.	La brisa „refrescándoles el cuerpo“ a los obreros (empatía táctil).
Perspectiva y	Filtro perceptivo	La narración refleja la	„Sintió cómo su

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

Categoría	Subcategoría	Descripción/ Operación	Evidencia del Cuento „El viaje de la brisa“
Subjetividad	del narrador	sensibilidad única del personaje.	alma quería liberarse“ (brisa percibiendo la muerte).
	Alteraciones perceptivas	Estados sensibles modifican la realidad narrada.	„Se tornó fría, atrajo nubarrones“ (tristeza transformando el clima).
Ritmo y Textura Narrativa	Sintaxis sensorial	Ritmo discursivo que imita sensaciones (lento/rápido, suave/áspero).	„Movimiento ondulante“ (frases fluidas que evocan el vaivén de la brisa).
	Densidad sensorial	Concentración de estímulos para intensificar escenas.	„Gotas de sudor caían de su frente... el niño lloró“ (clímax táctil y auditivo).

Claves Teóricas:

1. Semiótica de las pasiones (Deleuze): Los estados sensibles operan como „regímenes de signos“ que organizan la percepción y la emoción en el texto.
2. Inmersión: La brisa actúa como un „cuerpo sin órganos“ (Deleuze-Guattari), canalizando sensaciones que disuelven límites entre sujeto y entorno.
3. Ejemplo destacado: La escena del parto („entró por su nariz... el niño lloró“) muestra cómo lo sensorial construye pathos y ética narrativa.

Nota: Las evidencias del cuento se seleccionaron para ilustrar cómo cada categoría se materializa en la narración poética y sensorial de la brisa. ¿Necesitas ajustar enfoques teóricos o añadir más ejemplos?

Nota: Categorías de Análisis (basadas en la semiótica de las pasiones y los documentos proporcionados) como salida al *prompt* Gemini 2.0

Tabla 3

Análisis de Datos por Categorías: Semiótica de las Pasiones en la Configuración Narrativa

Pregunta de Investigación: ¿Cómo operan los estados sensibles en la configuración de la narración?

Categoría	Subcategoría	Descripción/ Operación	Evidencia del Cuento
Mediación Perceptiva	Selección de detalles	La brisa destaca aspectos de la realidad según su estado emocional o cognitivo.	„Se paseó entre los bañistas de una playa azul turquesa, jugó con unos niños que se reían“. La alegría enfoca en juegos y colores vivos.
	Jerarquización sensorial	Priorización de estímulos sensoriales vinculados al estado sensible.	„Recogió frío de una montaña y lo bajó hasta una gran ciudad, haciendo que sus habitantes se estremecieran“. El tacto (frío) domina la escena.
Modulación Discursiva	Ritmo sintáctico	Frases fluidas y ondulantes para reflejar el movimiento de la brisa.	„La brisita volaba sin cesar, a veces cansada se unía a otras brisas... “. Ritmo suave y continuo.
	Léxico y campos semánticos	Uso de palabras con carga afectiva („ligera“, „divertir“, „dolor“).	„Emocionada con las olas, imitaba su movimiento ondulante“. Campo semántico de alegría y dinamismo.
Temporalidad Narrativa	Duración percibida	El tiempo se dilata o comprime según la intensidad emocional.	„El niño lloró por primera vez y todos se sintieron aliviados“. Momento clave ralentizado.
	Secuencialidad	Transiciones entre escenas marcadas por cambios emocionales.	„Partió de allí hacia otros pueblos... “. La brisa cambia de lugar según su curiosidad.

Valoración crítica de la interacción Humano-Inteligencia Artificial (IA)

Categoría	Subcategoría	Descripción/ Operación	Evidencia del Cuento
Focalización y Perspectiva	Distancia perceptiva	La brisa alterna entre observación distante e inmersión íntima.	„Se volvió pequeña, casi imperceptible para acercarse al enfermo“. Acercamiento físico y emocional.
	Puntos ciegos narrativos	Omisión de detalles irrelevantes para el estado sensible.	La brisa ignora el entorno cuando se concentra en salvar al bebé: „no hubo tiempo de llegar al hospital“.
Estados Alterados	Subjetividad vs. objetividad	La brisa experimenta realidades híbridas (naturaleza y humanidad).	„Sintió cómo su alma quería liberarse...“. Fusión de percepción física y espiritual.
Función Emocional	Conexión con el lector	Empatía generada mediante acciones bondadosas de la brisa.	„Refrescándoles el cuerpo y animando su espíritu“ (obreros).
	Símbolos y tropos	Objetos neutrales adquieren significado afectivo.	„El pequeño papel con el número de teléfono“ simboliza esperanza para el chico tímido.

Ajustes Clave:

1. Evidencias específicas: Se reemplazaron ejemplos teóricos por fragmentos textuales del cuento.
2. Coherencia narrativa: Las categorías se adaptaron a la voz poética y perspectiva única de la brisa.
3. Semiótica de las pasiones: Se enfatiza cómo los estados sensibles (alegría, empatía, dolor) configuran la percepción y acción de la brisa.

Nota: Categorías de Análisis (basadas en la semiótica de las pasiones y el cuento „El viaje de la brisa“) como salida al *prompt* Claude Sonnet 4

Tabla 4

Análisis de Datos por Categorías: Semiótica de las Pasiones en la Configuración Narrativa

Pregunta de Investigación: ¿Cómo operan los estados sensibles en la configuración de la narración?

Categoría	Subcategoría	Descripción/Operación	Evidencia del Cuento
Modulación del Tono	Atmósfera emocional	Estados sensibles dominantes que impregnan el relato (euforia, melancolía, tensión).	„La brisa se tornó fría, atrajo nubarrones“ (tristeza que transforma el clima).
	Simbología ambiental	Elementos naturales asociados a emociones (lluvia=dolor, sol=alegría).	„Jugó con unos niños que se reían“ (luz/calor vinculados a felicidad).
Psicología de Personajes	Estados internos como motor	Decisiones y acciones guiadas por percepciones afectivas.	„Entró por su nariz... el niño lloró“ (compasión como impulso).
	Epifanías sensoriales	Revelaciones ligadas a experiencias corporales o emocionales.	„Comprendió que quien sonríe es el alma“ (epifanía en la muerte).
Estructura Simbólica	Motivos recurrentes	Objetos/sensaciones que adquieren significado afectivo (viento=libertad).	„El pequeño papel con el número“ (objeto que simboliza esperanza).
	Metáforas corporales	Fusión de lo físico y lo emocional (frío=tristeza, calor=empatía).	„Refrescándoles el cuerpo [a los obreros]“ (tacto como acto solidario).
Punto de Vista Sensible	Subjetividad radical	Narración filtrada por la percepción no humana de la brisa.	„Sintió cómo su alma quería liberarse“ (experiencia liminal entre vida/muerte).
	Alteraciones perceptivas	Realidad distorsionada por estados emocionales intensos.	„Se volvió pequeña, casi imperceptible“ (distanciamiento en el dolor).
Dinámica con el Lector	Empatía multisensorial	Conexión afectiva mediante estímulos sensoriales compartidos.	„El olor dulce a clavos y canela“ (olfato que evoca calidez doméstica).
	Política de lo sensible	Visibilización de experiencias marginales (parto, muerte, enfermedad).	„Gotas de sudor... el niño nació“ (escena de parto como acto político).

Claves Teóricas:

1. Semiótica de las pasiones (Deleuze): La brisa opera como „máquina deseante“ que conecta cuerpos, espacios y afectos.
2. Subjetividad no humana: La perspectiva de la brisa desestabiliza antropocentrismos narrativos.
3. Ejemplo destacado: La anciana que „dirige“ la brisa ilustra cómo lo sensible construye comunidades afectivas.

Nota: Las categorías integran lo teórico (modos semióticos) y lo textual (análisis del cuento). ¿Necesitas ajustar enfoques o profundizar en algún aspecto?

Nota: Categorías de Análisis (basadas en la semiótica de las pasiones y los documentos proporcionados) como salida al *prompt* DeepSeek.

3. A modo de cierre

Este ejercicio que combina la reflexión sobre puntos de vista expertos disciplinares y la interacción con la máquina resalta que el problema central sobre lo que se está presentando como una revolución tecnológica sin precedentes, no radica tanto en decidir si debemos rechazar o abrazar a los sistemas de „Inteligencia Artificial“ como herramienta destructiva, sino en cómo podemos desarrollar formas de interacción en las que los usuarios seamos conscientes tanto del potencial como de las limitaciones constitutivas de los modelos. Puesto que en este punto, casi irreversible de integración humano-máquina, no podemos asumir una postura definitiva sobre la „Inteligencia Artificial“, sino pensar en las condiciones bajo las cuales la interacción puede llegar a ser productiva, ética y semióticamente significativa para los seres humanos.

Al respecto podemos seguir la idea de Juan Magariños de Morentin (2009) de que todo sistema semiótico opera en sus propios bordes, y es justamente en ese borde donde se sitúan los sistemas de „Inteligencia Artificial“. Situados allí en el borde podemos comenzar a comprender el papel de estas herramientas no como sustitutas del pensamiento, la sensibilidad o la creatividad humana, sino como dispositivos de confrontación y revelación. En efecto, el temor, el miedo, el pánico, pasiones del sujeto, se manifiestan de manera necesaria puesto que nos enfrentamos a sistemas capaces de simular interpretación, articular sentido y producir coherencia lingüística con una velocidad y precisión impresionantes, pero sin cuerpo, sin historia, sin experiencia vivida y sentida. En este borde, el desafío no es técnico, sino hermenéutico y desde luego semiótico, por ser relaciones que conllevan a la acción.

Los análisis generados por los sistemas de „Inteligencia Artificial“ **consultados en este ejercicio de exploración** son la evidencia de que los sistemas muestran comprensión estructural a partir de la capacidad de imitar los marcos teóricos de la Semiótica de las Pasiones, por ejemplo. Y estas **observaciones que, definitivamente, son preliminares sugieren que** no captan de modo eficiente matices y sutilezas de la dimensión afectiva encarnada y las transformaciones pasionales del sujeto sintiente expuestas en sentido narrativo.

Es notable el potencial para clasificar, resumir, categorizar y formalizar, pero **el ejercicio resalta las** limitaciones cuando se trata de comprender lo implícito, la experiencia centrada en la pasión que se ficcionaliza en el texto; **aunque esto, está claro, requiere una verificación empírica más rigurosa que debe ir a la par del desarrollo de los sistemas, sobre todo si tomamos en cuenta las declaraciones de Geoffrey Hinton (2025) en las que señala que la „Inteligencia Artificial“ podría desarrollar emociones „in terms of cognitive behavior, it could be just like us when it comes to emotions“².**

En el campo de las ciencias humanas y ciencias sociales, es sumamente tentadora la fluidez del discurso automático y por esto resulta necesario interrogar los procesos lógicos de la máquina y rastrear la procedencia de los resultados, para poder evaluar los efectos en sectores sensibles como el educativo y el creativo; teniendo presente que los sistemas **parecen** mapear regularidades en el uso del lenguaje, pero **según lo que se presenta en este documento**, no pueden advertir por sí mismos las tensiones internas de los marcos teóricos que articulan, ni las consecuencias sociales de sus respuestas, al menos hasta que se verifique lo contrario. En este sentido el pensamiento crítico que va desde Foucault hasta Pasquinelli; desde Greimas hasta Chomsky, con sus distancias, ponen el ojo sobre la forma en que la tecnología se inserta en la política, el consumo y la producción de conocimiento. Puesto que no es sólo una cuestión de eficiencia o innovación, sino una cuestión de poder.

Este punto, en este estar en el borde de un sistema semiótico, nos recuerda que el sentido se construye en el encuentro con el otro, en la experiencia sensible, en la posibilidad de lo imprevisible. No se trata de competir con la máquina, ni de resistirla como enemigo, sino de aprender a dialogar con ella desde nuestras propias capacidades y límites, asumiendo asimismo, que la máquina también tiene límites.

² **Nota de la Autora:** En términos de „comportamiento cognitivo, podría ser igual que nosotros en lo que respecta a las emociones“ Traducción de Google Translate (Trabajo original publicado el 29 de Mayo, 2025 en la Red Social X, cuenta oficial @vitrupo de Adam McCann orientada a contenido sobre Inteligencia Artificial)

De este modo tanto el sujeto, la persona humana como la máquina **enfrentan el desafío** de asumir que la construcción del sentido es una tarea que requiere tanto de inteligencia como de cuerpo sintiente, finitud más que obsolescencia, memoria y deseo; **aunque el reconocimiento conceptual de esta premisa es apenas el punto de partida, pues la cuestión sustancial radica en cómo traducir esta conciencia teórica en prácticas concretas y en evidencia tangible, verificada.**³

Referencias

- Anthropic** (2025). Claude Sonnet 4 [Modelo de lenguaje de gran escala]. Disponible en: <https://claude.ai>
- Ávila, Alex** (2025, mayo). „AI wars: How corporations hijacked anti-AI backlash“ [Video]. YouTube. Disponible en: <https://www.youtube.com/watch?v=IRq0pESKJgg>
- Berger, Peter** y **Luckmann, Thomas** (1995). La construcción social de la realidad. Amorrortu.
- DeepSeek** (2025). DeepSeek [Modelo de lenguaje de gran escala]. Disponible en: <https://www.deepseek.com>
- Eco, Umberto** (1992). Los límites de la interpretación. Lumen.
- Fabrizi, Paolo** (2000). El giro semiótico. Gedisa.
- Foucault, Michel** (1978). Microfísica del poder (J. Varela & F. Alvarez-Uría, Trads.). Las Ediciones De.
- García, José** (s.f.). Manual de semiótica: Semiótica greimasiana narrativa, con aplicaciones y ejemplos en medios masivos de comunicación. Universidad de Lima.
- Google** (2025). Gemini 2.0 [Modelo de lenguaje de gran escala]. Disponible en: <https://gemini.google.com>
- Greengard, Samuel** (2025, mayo). „Shining a light on AI hallucinations: How they come about, their impact, and what can be done to mitigate or entirely end AI's tendency to make things up“. Communications of the ACM, 68(5). Disponible en: <https://doi.org/10.1145/3715691>
- Greimas, Algirdas** (1971). Semántica estructural: Investigación metodológica. Gredos.
- Greimas, Algirdas** (1973). En torno al sentido: Ensayos semióticos. Fragua.
- Greimas, Algirdas** y **Fontanille, Jacques** (1994). Semiótica de las pasiones: De los estados de cosas a los estados de ánimo. Siglo XXI.
- Hinton, Geoffrey** (2025, 29 de mayo). [Entrevista]. En **A. McCann** [@vitrupe], X (antes Twitter). Disponible en: <https://x.com/vitrupe>

³ **Nota de la Autora:** Para el desarrollo de este documento se utilizaron herramientas de Inteligencia Artificial con fines estrictamente académicos de revisión, consulta e interacción sistemática. El uso de estas herramientas tecnológicas se manteniendo la autoría intelectual del trabajo y cumpliendo con los estándares académicos de integridad investigativa. Todas las herramientas de IA utilizadas se citadas en la bibliografía

- Landowski, Erick** (2017). Presencias del otro. Universidad de Lima, Fondo Editorial.
- Lucifora, María Clara** (2021). „Cooperación semiótica entre los seres humanos y la Inteligencia Artificial“. Revista Internacional de Humanidades, 9(1), 17-30. Disponible en: <https://doi.org/10.18848/2474-5022/CGP/v09i01/17-30>
- Magariños, Juan** (2009). La semiótica de los bordes: Apuntes de metodología semiótica. Disponible en: <http://www.magariños.com.ar/Indice-Manual.html>
- McMahon, Liv** (2025, 23 de mayo). „AI system resorts to blackmail if told it will be removed“. BBC News. Disponible en: <https://www.bbc.com/news/articles/cpqeng9d20go>
- OpenAI** (2025). ChatGPT-4 [Modelo de lenguaje de gran escala]. Disponible en: <https://chat.openai.com>
- Pasquinelli, Matteo** (2022). „Cómo una máquina aprende y falla: Una gramatura del error para la Inteligencia Artificial“ (E. Cafassi, C. Monti, H. Peckaitis & G. Zarauza, Trads.). Revista Hipertextos, 10(17), 13-29. Disponible en: <https://doi.org/10.24215/23143924e054>
- Pasquinelli, Matteo** (2023). The eye of the master: A social history of artificial intelligence. Verso.
- Pizarro-Romero, Javier y Lovón, Marco** (2025). „El uso de la IA en cursos de redacción e investigación universitaria en el aula: Una experiencia de caso“. Desde el Sur, 17(1), e0015.
- Putnam, Hilary** (1990). Representación y realidad: Un balance crítico del funcionalismo. Gedisa.
- Robinson, Erhardt** [Anfitrión] (2025). „Ned Block on consciousness, artificial intelligence, and the border between seeing and thinking“ [Video]. YouTube. Disponible en: <https://www.youtube.com/watch?v=wM1fcZr0iSk>
- Theobald, Oliver** (2024). ChatGPT prompts book: Precision prompts, priming, training & AI writing techniques for mortals.
- Verón, Eliseo** (1993). La semiosis social: Fragmentos de una teoría de la discursividad. Gedisa.

Producción científica sobre alfabetización en inteligencia artificial: estudio bibliométrico en Web of Science.

Scientific production on artificial intelligence literacy: a bibliometric study in Web of Science.

Alberto Carballo Soca
alberto.carballo@psico.uh.cu

Resumen

Este estudio bibliométrico analiza la producción científica sobre alfabetización en inteligencia artificial (IA) registrada en la Web of Science hasta mayo de 2025. La investigación, que analizó 268 documentos, evidencia un crecimiento exponencial del interés académico en el tema, que permaneció latente durante años para luego despegar notablemente a partir de 2019, culminando en un auge significativo de publicaciones entre 2023 y 2024. Las publicaciones se concentran predominantemente en un núcleo selecto de revistas académicas, con Education and Information Technologies emergiendo como la fuente más prolífica. Un pequeño grupo de autores altamente productivos lidera la investigación, mostrando un impacto considerable a través de sus citas, aunque la vasta mayoría de los investigadores contribuye de manera más esporádica, un patrón característico de campos emergentes. Instituciones de Asia oriental, junto con reconocidas universidades norteamericanas y europeas, son las más productivas, con China y Estados Unidos dominando tanto el volumen de producción científica como las redes de colaboración internacional. El área de educación e investigación educativa se consolida como el eje central de la investigación, interconectando diversas disciplinas. Los términos inteligencia artificial y alfabetización en inteligencia artificial son las palabras claves más frecuentes, subrayando la importancia de desarrollar marcos educativos sólidos para articular la alfabetización en IA de manera efectiva.

Palabras claves: bibliométrico, producción, científica, alfabetización, inteligencia artificial.

Abstract

This bibliometric study analyzes the scientific production on artificial intelligence (AI) literacy registered in the Web of Science until May 2025. The research, which analyzed 268 papers, evidences an exponential growth of scholarly interest in the topic, which lay latent for years before taking off markedly from 2019, culminating in a significant boom in publications between 2023 and 2024. Publications are predominantly concentrated in a select core of academic journals, with Education and Information Technologies emerging as the most prolific source. A small group of highly productive authors leads the research, showing considerable impact through their citations, although the vast majority of researchers contribute more sporadically, a pattern characteristic of emerging fields. East Asian institutions, together with renowned North American and European universities, are the most productive, with China and the United States dominating both the volume of scientific output and international collaborative networks. The area of education and educational research is consolidating as the central axis of research, interconnecting various disciplines. The terms artificial intelligence and artificial intelligence literacy are the most frequent keywords, underlining the importance of developing solid educational frameworks to articulate AI literacy effectively.

Keywords: bibliometric, scientific production, literacy, artificial intelligence.

1. Introducción

La alfabetización en inteligencia artificial (IA) es un concepto emergente que describe el conjunto de conocimientos, habilidades y actitudes necesarios para comprender, usar y evaluar tecnologías de IA de manera efectiva, ética y crítica (Wang et al., 2022). Incluye aspectos técnicos, éticos, sociales y prácticos, y se considera fundamental para desenvolverse en una sociedad cada vez más influida por la IA. La alfabetización en IA abarca la comprensión de qué es la IA, cómo funciona, cómo reconocerla en la vida cotidiana, y cómo interactuar con sistemas de IA de manera informada y segura (Chiu et al., 2024; Pinski y Benlian, 2024; Stolpe y Hallström, 2024; Yim, 2024). Incluye capacidades tecnológicas, laborales, de interacción humano-máquina y de aprendizaje continuo (Cetindamar et al., 2022). Es relevante en la educación (desde primaria hasta adultos), en el entorno laboral y en la vida cotidiana, adaptándose a las necesidades de diferentes grupos de usuarios (Laupichler et al., 2022; Sperling et al., 2024).

Por otro lado, realizar un estudio bibliométrico es fundamental para analizar de manera objetiva y sistemática la producción científica en un campo específico. Permite identificar patrones, tendencias, impacto y evolución de la investigación, facilitando la toma de decisiones informadas y el desarrollo teórico y práctico en distintas disciplinas. Los estudios bibliométricos ayudan a visualizar cómo evoluciona un campo, detectando áreas emergentes, temas en crecimiento y cambios en la producción científica (Donthu et al., 2021; Ellegaard y Wallin, 2015; Jing et al., 2023; Khan et al., 2021; Mukherjee et al., 2022).

Permiten medir la cantidad, calidad e impacto de publicaciones, autores, instituciones y países, siendo clave para la evaluación de la investigación y la toma de decisiones en políticas científicas (Ellegaard y Wallin, 2015; Passas, 2024; Salinas-Ríos y López, 2022). Facilitan la construcción de marcos teóricos integrados y contribuyen a la mejora de la práctica profesional al vincular resultados analíticos con el avance del conocimiento (Brika et al., 2021; Mukherjee et al., 2022; Zupic y Čater, 2014).

Las técnicas de mapeo bibliométrico permiten representar gráficamente la estructura intelectual y las relaciones entre temas, autores o instituciones (Jing et al., 2023; Zupic y Čater, 2014). Introducen objetividad en la revisión de literatura, reduciendo el sesgo del investigador al basarse en datos cuantitativos y análisis estadísticos (Passas, 2024; Salinas-Ríos y López, 2022; Zupic y Čater, 2014). Son más accesibles y prácticos que revisiones sistemáticas o meta-análisis, requiriendo menos recursos y permitiendo su aplicación en la mayoría de las áreas científicas (Passas, 2024; Salinas-Ríos y López, 2022). Sirven como punto de partida para revisiones sistemáticas y meta-análisis, y ayudan a gestionar el conocimiento basado en evidencia (Donthu et al., 2021; Salinas-Ríos y López, 2022).

Realizar un estudio bibliométrico sobre alfabetización en IA se justifica por el crecimiento, la fragmentación y la interdisciplinariedad de la investigación en este ámbito, así como por la necesidad de mapear tendencias, identificar vacíos y orientar futuras agendas de investigación y políticas educativas (Jantakun et al., 2025). Permite identificar tendencias de publicación, redes de colaboración entre autores e instituciones, y temas emergentes como la integración tecnológica, la ética y la personalización del aprendizaje (Jantakun et al., 2025; Son y Jho, 2024; Wu et al., 2024). Ayuda a clarificar la relación y solapamiento entre alfabetización en IA y otras categorías, facilitando la integración conceptual y curricular (Chee et al., 2024; Son y Jho, 2024). Proporciona una base objetiva para diseñar políticas educativas, desarrollar programas de formación y orientar futuras investigaciones, asegurando que la alfabetización en IA responda a las necesidades reales de la sociedad (Chee et al., 2024; Jantakun et al., 2025; Pinski y Benlian, 2024).

Producción científica sobre alfabetización en inteligencia artificial

Es por ello que el objetivo de esta investigación es analizar la producción científica sobre alfabetización en IA indexada en la Web of Science (WoS) entre 1992 y 2025, identificando tendencias temáticas, dinámicas de colaboración y fuentes más influyentes.

2. Metodología

Se seleccionó la Web of Science Core Collection (WoS) como fuente principal de datos. De acuerdo con la descripción oficial de la plataforma (Clarivate Analytics, s.f.), la WoS se posiciona como la base de citas líder a nivel global. Este repositorio incluye registros de artículos científicos procedentes de revistas de alto impacto internacional —incluyendo publicaciones de acceso abierto—, así como actas de conferencias y libros académicos. La cobertura temporal del sistema abarca registros históricos en ciertas disciplinas desde 1900, aunque el acceso específico a los materiales puede variar según los alcances de la suscripción institucional.

Teniendo en cuenta su accesibilidad, se seleccionaron las siguientes colecciones de la WoS:

- Science Citation Index Expanded (SCI-EXPANDED): índice multidisciplinar de la literatura científica, que incluye todas las referencias citadas de los artículos indexados.
- Social Sciences Citation Index (SSCI): índice multidisciplinar de la literatura en ciencias sociales, que también incorpora selectivamente publicaciones relevantes de las principales revistas científicas y técnicas.
- Arts & Humanities Citation Index (AHCI): índice multidisciplinar de la literatura en artes y humanidades, al que se suman artículos seleccionados de revistas destacadas en ciencia y ciencias sociales.

La estrategia de búsqueda se aplicó en el campo „Título“ y combinó mediante el operador OR los siguientes términos: „artificial intelligence literac“, „AI literac“, „artificial intelligence competenc“, „AI competenc“, „artificial intelligence education“, „AI education“, „artificial intelligence awareness“, „AI awareness“, „artificial intelligence readiness“, „AI readiness“, „teaching AI“, „teaching artificial intelligence“, „learning AI“, „learning artificial intelligence“, „AI curriculum“, „artificial intelligence curriculum“, „AI skills“, „artificial intelligence skills“, „AI ethics education“ y „artificial intelligence ethics education“. Se excluyeron expresamente „machine learning“ y „deep learning“ para evitar referencias alejadas del tema central. Además, se aplicó un filtro de tipo de documento para recuperar únicamente artículos originales y de revisión, sin imponer restricciones de idioma ni de periodo.

La consulta devolvió 268 registros (265 en inglés y 3 en alemán), con cobertura temporal desde 1992 hasta la fecha de búsqueda (5 de mayo de

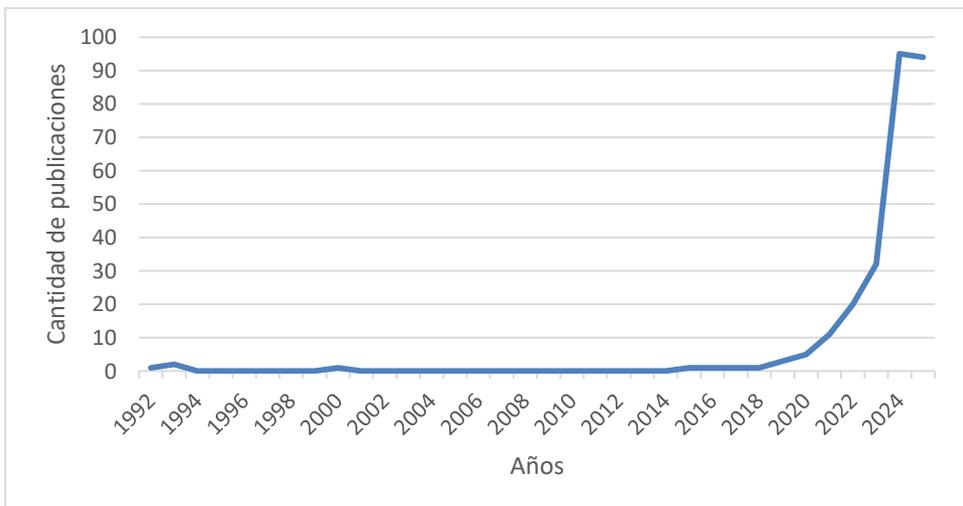
2025). Todos los registros se exportaron en formato de texto plano („Registro completo y referencias citadas“) e importaron en R mediante el paquete bibliometrix/biblioshiny (Aria y Cuccurullo, 2017), lo que permitió calcular métricas descriptivas, explorar interactivamente tendencias, generar gráficos, construir mapas de coocurrencia de palabras claves, clústeres temáticos y redes de colaboración.

3. Resultados

Descripción general

El análisis abarca el periodo 1992–2025 y considera 268 documentos extraídos de 144 fuentes académicas. El corpus presenta una tasa de crecimiento anual del 14,8%, con una edad media de publicaciones de 1,7 años y un promedio de 15,8 citas por documento. En total se incluyen 12 650 referencias y 793 palabras claves, escritas por 916 autores (de los cuales 23 son autores únicos de 27 documentos). El grado de colaboración es de 3,9 coautores por artículo y la proporción de colaboraciones internacionales alcanza el 26,5%. Como se muestra en la Figura 1, la investigación sobre alfabetización en IA fue de poco interés para la comunidad académica entre 1992 y 2019 (menos de 5 artículos anuales). A partir de 2019 comienza un ascenso moderado, que se dispara en 2023–2024, alcanzando 95 publicaciones en 2024 y 94 en 2025 hasta la fecha de corte del análisis. Este patrón refleja el progresivo interés académico en el tema.

Figura 1.
Producción anual de publicaciones sobre alfabetización en IA



Fuente: Elaboración propia.

Producción científica sobre alfabetización en inteligencia artificial

La Tabla 1 presenta la media de citas por artículo, por año para los artículos publicados anualmente, el número de artículos (N) publicados cada año y el número de años citables hasta la fecha en que se realizó el análisis. Aunque los artículos más antiguos han tenido más tiempo para acumular citas, el promedio de citas por año nos permite realizar una comparación más equitativa del impacto anualizado. Por ejemplo, los artículos publicados en 2017 y 2020 muestran un alto impacto promedio anualizado, lo que sugiere una recepción e influencia significativas por parte de la comunidad investigadora poco después de su publicación.

Tabla 1
Promedio de citas

Año	Promedio de citas por artículo	N	Promedio de citas por año	Años citables
1992	2.00	1	0.06	34
1993	2.00	2	0.06	33
2000	3.00	1	0.12	26
2015	21.00	1	1.91	11
2016	19.00	1	1.90	10
2017	146.00	1	16.22	9
2018	8.00	1	1.00	8
2019	26.33	3	3.76	7
2020	89.40	5	14.90	6
2021	75.09	11	15.02	5
2022	42.65	20	10.66	4
2023	25.97	32	8.66	3
2024	9.60	95	4.80	2
2025	0.86	94	0.86	1

Fuente: Elaboración propia.

Un análisis más exhaustivo del impacto se observa en los años más recientes, con un mayor número de publicaciones. Por ejemplo, en 2020, con cinco artículos, la media de citas por artículo fue de 89,40, lo que resultó en un promedio de citas por año de 14,90. Es interesante observar la tendencia en los años más cercanos (2022-2024), donde el número de publicaciones es considerablemente alto. Aunque el promedio de citas por artículo pueda ser menor debido al menor tiempo disponible para ser citados, el promedio de citas por año todavía puede indicar una fuerte captación inicial. Los datos de 2025, con 94 artículos y una media de citas por artículo de 0.86, reflejan publicaciones muy recientes con un tiempo de citación mínimo, lo cual es de esperar.

Fuentes

La Tabla 2 muestra las fuentes con mayor número de publicaciones sobre alfabetización en IA. La caracterización de estas revistas académicas revela un panorama diverso que refleja la amplitud de los campos de la educación, la tecnología, la medicina, la hostelería y las ciencias del comportamiento.

Tabla 2

Fuentes con mayor número de publicaciones sobre alfabetización en IA

Fuentes	Rango	Frecuencia	Frecuencia acumulada
Education and Information Technologies	1	33	33
Interactive Learning Environments	2	10	43
BMC Medical Education	3	8	51
IEEE Access	4	7	58
Sustainability	5	7	65
European Journal of Education	6	6	71
Applied Sciences	7	5	76
International Journal of Contemporary Hospitality Management	8	5	81
Academic Radiology	9	4	85
Behavioral Sciences	10	4	89

Fuente: Elaboración propia.

Producción científica sobre alfabetización en inteligencia artificial

La Zona 1, compuesta por 10 revistas, concentra el 33.2% de los 268 artículos analizados, validando la Ley de Bradford (1934) al demostrar que un núcleo reducido de fuentes genera un tercio de la producción científica relevante. Destaca Education and Information Technologies (33 artículos), que por sí sola aporta el 12.3% del total, seguida por revistas como Interactive Learning Environments (10) y BMC Medical Education (8), junto a títulos interdisciplinarios como IEEE Access y Sustainability (7 cada una), lo que sugiere una convergencia temática entre tecnología, educación y sostenibilidad. La rápida acumulación de artículos (43% del total en las primeras 5 revistas) subraya la eficiencia de priorizar este núcleo para acceder a la literatura clave.

Autoría

Los resultados obtenidos revelan patrones claros en la dinámica de autoría. La Tabla 3, dedicada a los autores más prolíficos, muestra que un pequeño conjunto de investigadores lidera la producción. Esta diferencia entre recuento bruto y valor fraccionalizado apunta a una alta colaboración, ya que ninguno de ellos publica en solitario la mayoría de sus trabajos.

Tabla 3

Autores más relevantes sobre alfabetización en IA

Autores	Artículos	Artículos fraccionalizados
Su JH	11	5.70
Chiu TKF	9	2.28
Chai CS	8	1.76
Dai Y	7	3.06
Jong MSY	6	1.43
Kim K	6	2.83
Chu SKW	5	1.28
Kim J	5	1.75
Kwon K	5	2.33
Ng DTK	5	1.37

Fuente: Elaboración propia.

El análisis de impacto (Tabla 4) confirma la relevancia de estos autores: Chai CS y Chiu TKF muestran un índice h de 7, un índice g de 8–9 y más de 580 citas totales cada uno; Su JH, con un índice h de 6 e índice g de 11, acumula

142 citas, a pesar de haber comenzado a publicar sólo desde 2020. Estos datos evidencian no sólo productividad, sino también creciente reconocimiento de su obra en los últimos cinco años.

Tabla 4

Análisis de impacto de los autores

Autor	Índice H	Índice G	Índice M	Citas totales	Cantidad de publicaciones	Año de inicio de las publicaciones
Chai CS	7	8	1.167	582	8	2020
Chiu TKF	7	9	1.167	581	9	2020
Su JH	6	11	3	142	11	2024
Chu SKW	5	5	2.5	104	5	2024
Dai Y	5	7	1	324	7	2021
Jong MSY	5	6	1	221	6	2021
Ng DTK	5	5	2.5	103	5	2024
Kim K	3	4	1.5	21	6	2024
Kong SC	3	3	1	84	3	2023
Kwon K	3	4	1.5	18	5	2024
Leung JKL	3	3	1.5	62	3	2024
Yang WP	3	4	1.5	64	4	2024

Fuente: Elaboración propia.

La aplicación de la Ley de Lotka (1926), como se aprecia en la Tabla 5, muestra que el 91,9% de los 916 autores contribuyó con un único documento, un 5,6% con dos, y sólo el 0,1% superó los cuatro trabajos.

Este patrón de muchos autores ocasionales versus pocos autores muy activos es característico de campos emergentes, donde un núcleo reducido de especialistas impulsa la continuidad del área mientras numerosos investigadores participan de forma puntual.

Tabla 5

Ley de Lotka

Documentos escritos	Número de autores	Proporción de autores
1	842	0.919
2	51	0.056
3	12	0.013
4	1	0.001

Producción científica sobre alfabetización en inteligencia artificial

5	4	0.004
6	2	0.002
7	1	0.001
8	1	0.001
9	1	0.001
11	1	0.001

Fuente: Elaboración propia.

La Tabla 6 sintetiza las instituciones más productivas en el campo de la alfabetización en IA. Predominan instituciones de Asia oriental, en paralelo con la presencia consolidada de universidades norteamericanas y europeas. El hecho de que solo cuatro instituciones superen las diez publicaciones destaca la existencia de un núcleo institucional relativamente reducido que lidera la investigación, mientras que el elevado número de afiliaciones con uno o dos documentos (más de 450 en total) apunta a una comunidad global ampliamente dispersa.

Tabla 6

Instituciones más relevantes en publicaciones sobre alfabetización en IA

Afiliación	Artículos
Universidad China de Hong Kong	35
Universidad de Educación de Hong Kong	16
Universidad de Corea	16
Universidad de Hong Kong	15
Universidad de Toronto	12
Universidad Normal de China Oriental	10
Universidad Nacional de Singapur	9
Universidad de Bonn	9
Universidad de Stanford	8
Universidad de Harvard	7
Universidad Normal del Sur de China	7
Universidad de Calgary	7
Banco Egipcio de Conocimientos	6

Fuente: Elaboración propia.

El análisis de la producción científica por países revela patrones distintivos en cuanto a la contribución y colaboración por países, los cuales se resumen en la Tabla 7. Los resultados indican una concentración significativa de la producción científica en un grupo selecto de países. China emerge como el líder en volumen de investigación, seguido por EE.UU. Conjuntamente, estos dos países aglutinan más de la mitad de la investigación publicada.

Tabla 7

Países más relevantes en publicaciones sobre alfabetización en IA

País	Artículos	Artículos (%)	Publicaciones de un solo país	Publicaciones multinacionales	Publicaciones multinacionales (%)
China	105	39.2	82	23	21.9
EE.UU.	37	13.8	27	10	27
Corea	24	9	23	1	4.2
Alemania	11	4.1	9	2	18.2
Reino Unido	8	3	3	5	62.5
Finlandia	6	2.2	2	4	66.7
Australia	5	1.9	4	1	20
Canadá	5	1.9	4	1	20
Irán	5	1.9	3	2	40
Arabia Saudita	4	1.5	1	3	75

Fuente: Elaboración propia.

En lo referente a las dinámicas de colaboración, la Tabla 7 muestra una variada aproximación entre los países más productivos. Por ejemplo, mientras China presenta un considerable número de publicaciones de un solo país, indicativo de una robusta capacidad investigadora interna, también participa activamente en la colaboración internacional. Por otro lado, países como Reino Unido y Finlandia exhiben un alto índice de colaboración internacional, lo que sugiere una fuerte integración en redes de investigación globales y una posible estrategia de complementación de capacidades a través de la cooperación transnacional.

Documentos

Las Tablas 8 y 9 recopilan los documentos con mayor impacto local dentro de la base de datos analizada, pero bajo métricas complementarias. La Tabla 8 prioriza aquellos con el mayor volumen absoluto de citas, reflejando su influencia bruta en la comunidad. Por su parte, la Tabla 9 emplea el promedio de citas por año, un indicador ajustado que nivela las diferencias temporales y ofrece una valoración más equitativa del impacto sostenido.

Producción científica sobre alfabetización en inteligencia artificial

Tabla 8

Documentos más citados localmente

Documento	Año	Citas
Wang BC, 2023, Behaviour & Information Technology	2023	37
Casal-Otero L, 2023, International Journal of STEM Education	2023	29
Chai CS, 2021, Educational Technology & Society	2021	25
Karaca O, 2021, BMC Medical Education	2021	22
Chiu TKF, 2022, IEEE Transactions on Education	2022	22
Kong HY, 2021, International Journal of Contemporary Hospitality Management	2021	20
Chiu TKF, 2020, Sustainability	2020	19
Liang XD, 2022, Tourism Management	2022	19
Su JH, 2024, Interactive Learning Environments	2024	16
Cetindamar D, 2024, IEEE Transactions on Engineering Management	2024	16

Fuente: Elaboración propia.

Tabla 9

Documentos más citados localmente (normalizado)

Documento	Año	Total de citas normalizado
El-Sayed BKM, 2025, Nurse Education in Practice	2025	10.85
Wang XC, 2025, European Journal of Education	2025	10.85
Kim K, 2025, Interact Learn Envir	2025	10.85
Wang BC, 2023, Behav Inform Technol	2023	8.05
Su JH, 2024, Interactive Learning Environments	2024	7.92
Cetindamar D, 2024, IEEE Transactions on Engineering Management	2024	7.92
Kong SC, 2025, Innovations in Education and Teaching International	2025	7.23
Casal-Otero L, 2023, International Journal of STEM Education	2023	6.31
Chiu TKF, 2020, Sustainability	2020	5.00

Ng DTK, 2024, Interactive Learning Environments	2024	4.95
Teng RR, 2024, International Journal of Contemporary Hospitality Management	2024	4.95

Fuente: Elaboración propia.

La Tabla 8 muestra que los documentos con mayor impacto bruto corresponden a publicaciones entre 2020 y 2024. Destacan las contribuciones de Wang BC (2023) y Casal-Otero L (2023), lo que indica una amplia referencia a sus trabajos dentro de la comunidad analizada. La presencia recurrente de autores como Chiu TKF (2020, 2022) sugiere una línea de investigación sostenida en el tiempo. Al normalizar las citas por año (Tabla 9), se observan ajustes significativos en el ranking de impacto. Los documentos más recientes presentan valores elevados de citas normalizadas, lo que refleja que, al controlar el factor temporal, las publicaciones recientes exhiben una tasa de adopción más acelerada, probablemente debido a la constante actualización del campo. Autores como Chiu TKF, Su JH, Wang BC, Cetindamar D y Casal-Otero L aparecen en ambas tablas, lo que indica que sus trabajos combinan impacto acumulado y relevancia reciente.

La Tabla 10 presenta las referencias de mayor impacto a nivel local, conformando un núcleo de estudios recientes (2020-2023). Se destacan publicaciones en revistas y conferencias especializadas, reflejando una evolución constante del campo a través de su distribución temática y temporal.

Tabla 10

Referencias más citadas localmente

Referencia citada	Citas
Long, D. R., 2020, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), DOI 10.1145/3313831.3376727	73
Ng, D. T. K., 2021, Computers and Education: Artificial Intelligence, v2, DOI 10.1016/j.caeai.2021.100041	55
Wang, B. C., 2023, Behaviour & Information Technology, v42, p1324, DOI 10.1080/0144929X.2022.2072768	37
Fornell, C., 1981, Journal of Marketing Research, v18, p39, DOI 10.2307/3151312	34
Casal-Otero, L., 2023, International Journal of STEM Education, v10, DOI 10.1186/s40594-023-00418-7	29
Kandlhofer, M., 2016, Proceedings of the Frontiers in Education Conference	29
Chai, C. S., 2021, Educational Technology & Society, v24, p89	25
Touretzky, D., 2019, Proceedings of the AAAI Conference on	25

Producción científica sobre alfabetización en inteligencia artificial

Artificial Intelligence, p9795	
Zhang, H., 2023, International Journal of Artificial Intelligence in Education, v33, p290, DOI 10.1007/s40593-022-00293-3	25
Kong, S.-C., 2021, Computers and Education: Artificial Intelligence, v2, DOI 10.1016/j.caeai.2021.100026	24

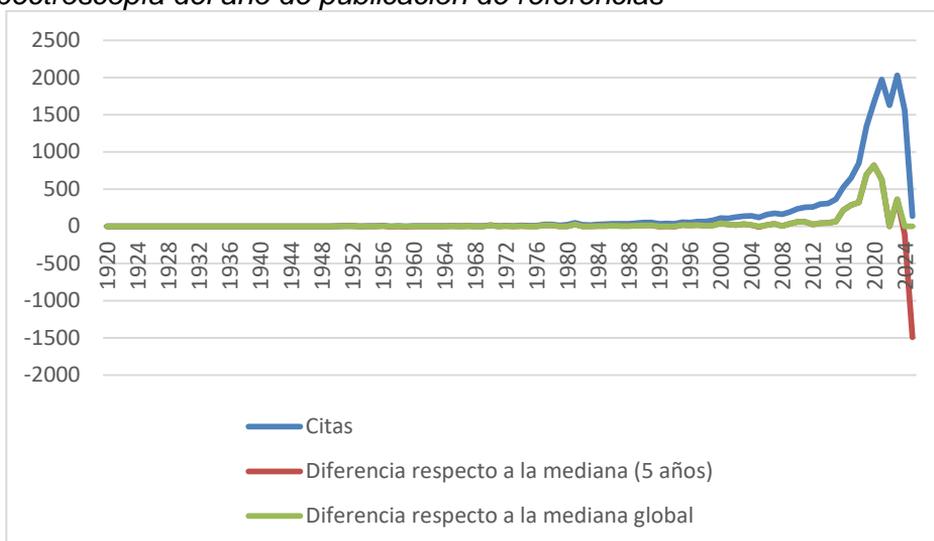
Fuente: Elaboración propia.

La Figura 2 revela una evolución histórica marcada por tres fases claras en la producción y citación de literatura relacionada con el tema de estudio:

1. Período de latencia (1920-1940): Casi nula actividad académica (0-1 cita/año), lo que sugiere que el tema aún no se había consolidado como área de investigación.
2. Crecimiento incipiente (1950-1990): Aumento gradual, con picos notables en 1977-1978 (24-25 citas) y un hito en 1981 (48 citas, más 24 respecto a la mediana).
3. Explosión temática (2000-2023): Crecimiento exponencial, especialmente desde 2016 (524 citas, más 24 respecto a la mediana), alcanzando un máximo en 2021 (1972 citas), lo que refleja la madurez del campo y su relevancia actual.

Figura 2.

Espectroscopía del año de publicación de referencias



Fuente: Elaboración propia.

La diferencia respecto a la mediana global destaca años clave como 1981, 2016-2021 y 2023, confirmando su influencia paradigmática. Este patrón

sugiere una disciplina joven, con raíces en el siglo XX pero consolidada en las últimas décadas, impulsada por avances tecnológicos y demandas sociales.

La Tabla 11 evidencia un ámbito de investigación prioritariamente centrado en educación. Paralelamente, se identifican líneas de interés secundario en estudios ambientales, contextos organizacionales, aplicaciones en salud y aproximaciones multidisciplinares, que amplían el espectro temático del campo analizado.

Tabla 11
Áreas de investigación más frecuentes

Áreas de investigación	Frecuencia
Educación e investigación educativa	106
Disciplinas científicas de la educación	27
Gestión	20
Negocios	18
Ingeniería multidisciplinaria	16
Hostelería, ocio, deporte y turismo	13
Ciencias de la información y bibliotecología	12
Ingeniería eléctrica y electrónica	11
Estudios ambientales	11
Psicología multidisciplinaria	11
Ciencias de la computación - Sistemas de información	10
Radiología, medicina nuclear e imágenes médicas	10

Fuente: Elaboración propia.

La Tabla 12 muestra que los términos inteligencia artificial y alfabetización en inteligencia artificial son las palabras claves más frecuentes en el corpus analizado, lo que indica su centralidad en la literatura sobre alfabetización en IA. Le siguen en relevancia conceptos asociados a contextos educativos, como educación en inteligencia artificial, educación superior y educación K-12, así como grupos específicos como estudiantes de medicina.

La presencia de términos técnicos como aprendizaje automático e inteligencia artificial generativa señala un vínculo explícito con herramientas tecnológicas concretas, mientras que autoeficacia y preparación para la inteligencia artificial reflejan dimensiones actitudinales y psicosociales recurrentes en los estudios.

Producción científica sobre alfabetización en inteligencia artificial

Tabla 12

Palabras claves más frecuentes

Palabras claves	Frecuencia
Inteligencia artificial (AI)	108
Alfabetización en inteligencia artificial (artificial intelligence literacy)	76
Educación en inteligencia artificial (artificial intelligence education)	36
Conciencia sobre inteligencia artificial (artificial intelligence awareness)	19
Preparación para la inteligencia artificial (artificial intelligence readiness)	17
Aprendizaje automático (machine learning)	16
Educación (education)	15
Inteligencia artificial generativa (generative artificial intelligence)	12
Educación superior (higher education)	11
Autoeficacia (self-efficacy)	11
Educación básica (k-12 education)	8
Estudiantes de medicina (medical students)	8

Fuente: Elaboración propia.

Nota: Los términos en español incluyen su equivalente en inglés para asegurar consistencia con la nomenclatura académica internacional y optimizar la recuperación de información en repositorios científicos.

Estructura conceptual

La Tabla 13 y la Figura 3 muestran los grupos temáticos a partir de las áreas de investigación. El grupo 5 emerge como un nodo crítico, liderado por educación e investigación educativa, con el valor más alto de intermediación y un PageRank elevado. Esto indica su rol central como puente interdisciplinario, integrando campos tan diversos como enfermería, lingüística y ciencias de la salud y servicios. Su predominio sugiere que los marcos educativos son fundamentales para articular la alfabetización en IA en contextos aplicados, trascendiendo barreras disciplinares.

Tabla 13
Nodos temáticos por área de investigación

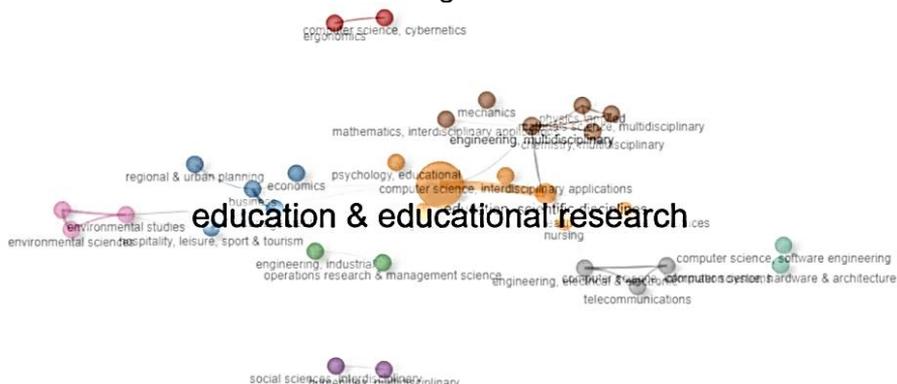
Nodo	Grupo	Intermediación	Cercanía	PageRank
Ciencias de la computación, cibernética	1	0	1	0.031
Ergonomía	1	0	1	0.031
Gestión	2	122	0.018	0.064
Negocios	2	47	0.016	0.049
Hostelería, ocio, deporte y turismo	2	0	0.013	0.029
Economía	2	0	0.012	0.010
Planificación regional y urbana	2	0	0.012	0.015
Ingeniería industrial	3	24	0.013	0.021
Investigación operativa y ciencia de la gestión	3	0	0.010	0.013
Ciencias sociales interdisciplinarias	4	0	1	0.031
Humanidades multidisciplinares	4	0	1	0.031
Educación e investigación educativa	5	189	0.021	0.079
Educación en disciplinas científicas	5	107	0.018	0.073
Enfermería	5	0	0.013	0.014
Aplicaciones interdisciplinarias en ciencias de la computación	5	0	0.014	0.012
Lingüística	5	0	0.014	0.012
Ciencias de la salud y servicios	5	0	0.013	0.009
Psicología educativa	5	0	0.014	0.009
Ingeniería multidisciplinaria	6	107	0.018	0.074
Química multidisciplinaria	6	0	0.013	0.035
Ciencia de materiales multidisciplinaria	6	0	0.013	0.035
Física aplicada	6	0	0.013	0.035
Aplicaciones interdisciplinarias en matemáticas	6	0	0.013	0.009
Mecánica	6	0	0.013	0.009
Estudios ambientales	7	46	0.013	0.047
Ciencias ambientales	7	0	0.010	0.031
Ciencia y tecnología verde y sostenible	7	0	0.010	0.031
Ingeniería eléctrica y electrónica	8	46	0.013	0.036
Sistemas de información en ciencias de la computación	8	0	0.010	0.031
Telecomunicaciones	8	0	0.010	0.031
Ingeniería de software	9	0	1	0.031
Arquitectura y hardware en ciencias de la computación	9	0	1	0.031

Fuente: Elaboración propia.

Producción científica sobre alfabetización en inteligencia artificial

Figura 3

Red de coocurrencia de áreas de investigación



Fuente: Elaboración propia.

En el grupo 2, gestión y negocios, destacan por su intermediación alta, lo que significa que conectan áreas técnicas con aplicaciones prácticas en organizaciones. Sin embargo, su baja cercanía sugiere que, aunque son puentes clave, tienen poca cohesión interna, lo que podría dificultar la integración de conceptos dentro de estos campos. Por otro lado, grupos como 1, 8 y 9 tienen cercanía máxima, lo que indica que son campos muy consolidados internamente. Pero su intermediación nula revela que, a pesar de su especialización, tienen poca influencia en la red global, posiblemente por enfocarse en nichos específicos. En contraste, en el grupo 6, ingeniería multidisciplinaria, y en el 7 estudios ambientales muestran métricas moderadas, lo que muestra su creciente papel en aplicar IA a retos técnicos y ecológicos. No obstante, su baja cercanía sugiere que necesitan más alianzas con áreas como educación o gestión para ampliar su impacto.

La Tabla 14 y la Figura 4 muestran los grupos temáticos a partir de las palabras clave. El grupo 1 es el conector central de la red, con alta intermediación y PageRank, integrando educación, ética y salud. Sin embargo, sus subtemas dependen casi exclusivamente del nodo IA, lo que revela una influencia concentrada. El grupo 3 destaca por vincular educación superior y sostenibilidad con herramientas como ChatGPT. Aunque tiene relevancia, su baja cercanía sugiere fragmentación interna por abordar diversos temas.

Tabla 14

nodos temáticos por palabras claves

Nodo	Grupo	Intermediación	Cercanía	PageRank
IA (artificial intelligence)	1	607.631	0.019	0.22
Preparación en IA (artificial intelligence readiness)	1	0.095	0.011	0.02

Aprendizaje automático (machine learning)	1	42	0.011	0.022
Educación (education)	1	2.697	0.012	0.036
Autoeficacia (self-efficacy)	1	2.293	0.012	0.02
Estudiantes de medicina (medical students)	1	1.095	0.012	0.025
Ética (ethics)	1	0	0.011	0.02
Alfabetización (literacy)	1	0	0.011	0.007
Desarrollo de escalas (scale development)	1	0.089	0.011	0.013
Educación médica (medical education)	1	0	0.011	0.018
Bibliografías (bibliographies)	1	0	0.011	0.014
Chatbot (chatbot)	1	0	0.011	0.01
Educación en enfermería (nursing education)	1	0	0.011	0.008
Validez y confiabilidad (validity and reliability)	1	0	0.011	0.013
Tecnología educativa (educational technology)	1	0	0.011	0.007
Empleo (employment)	1	0	0.011	0.007
Innovación (innovation)	1	0	0.011	0.006
Estudiantes de enfermería (nursing students)	1	0	0.011	0.007
Educación en IA (artificial intelligence education)	2	110.236	0.014	0.072
Educación básica (k-12 education)	2	5.239	0.012	0.023
Pensamiento computacional (computational thinking)	2	0	0.012	0.013
Formación docente (teacher education)	2	0.211	0.012	0.014
Ética en IA (ai ethics)	2	0	0.010	0.009
Currículo de IA (artificial intelligence curriculum)	2	0	0.009	0.006
Teoría de la autodeterminación (self-determination theory)	2	0	0.010	0.011
Revisión sistemática (systematic review)	2	0	0.009	0.009
Diseño curricular (curriculum design)	2	0	0.011	0.012
Alfabetización en IA (artificial intelligence literacy)	3	248.096	0.016	0.138

Producción científica sobre alfabetización en inteligencia artificial

IA generativa (generative artificial intelligence)	3	0	0.011	0.014
Educación superior (higher education)	3	0	0.011	0.018
IA en educación (ai in education)	3	0	0.009	0.007
Intención conductual (behavioral intention)	3	0	0.012	0.012
ChatGPT (chatgpt)	3	0	0.011	0.013
Educación infantil (early childhood education)	3	0.319	0.012	0.017
Pedagogía (pedagogy)	3	0	0.011	0.011
Currículo (curriculum)	3	0	0.011	0.011
Modelado de ecuaciones estructurales (structural equation modeling)	3	0	0.009	0.006
Desarrollo sostenible (sustainable development)	3	0	0.011	0.008
Escuela primaria (elementary school)	3	0	0.009	0.006
Diseño instruccional (instructional design)	3	0	0.011	0.008
Educación secundaria (secondary education)	4	0	1	0.022
Habilidades del siglo XXI (twenty-first century skills)	4	0	1	0.022
Conciencia sobre IA (artificial intelligence awareness)	5	83	0.011	0.025
Ciencia de datos (data science)	6	0	0.008	0.006
Agotamiento emocional (emotional exhaustion)	7	0	0.008	0.009
Inseguridad laboral (job insecurity)	8	0	0.008	0.009

Fuente: Elaboración propia.

Nota: Los términos en español incluyen su equivalente en inglés para asegurar consistencia con la nomenclatura académica internacional y optimizar la recuperación de información en repositorios científicos.

El grupo 2 conecta teoría y práctica (ej.: educación básica y diseño curricular), pero su baja cercanía refleja dificultades para consolidar enfoques unificados. En contraste, el grupo 4 tiene máxima cohesión interna, pero está aislado del resto. Los grupos del 5 al 8 son periféricos: temas como conciencia sobre IA o

Producción científica sobre alfabetización en inteligencia artificial

Journal of Management	1	14.616	0.015	0.017
Journal of Vocational Behaviour	1	0.402	0.012	0.012
Journal of Retailing and Consumer Services	1	3.742	0.014	0.01
Computers & Education: Artificial Intelligence	2	1.702	0.012	0.045
Education and Information Technologies	2	1.597	0.012	0.044
Computers & Education	2	2.32	0.012	0.04
Sustainability	2	5.062	0.012	0.039
Computers in Human Behavior	2	12.699	0.012	0.035
arXiv	2	0.663	0.012	0.016
AAAI Conference on Artificial Intelligence	2	0.25	0.011	0.028
Interactive Learning Environments	2	0.871	0.012	0.033
British Journal of Educational Technology	2	0.819	0.012	0.03
International Journal of Artificial Intelligence in Education	2	0.475	0.012	0.031
Educational Technology & Society	2	0.801	0.012	0.031
Frontiers in Psychology	2	2.709	0.012	0.022
International Journal of Educational Technology in Higher Education	2	0.647	0.012	0.023
Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems	2	1.412	0.012	0.033
Educational Technology Research and Development	2	0.303	0.011	0.023
IEEE Access	2	0.202	0.011	0.021
Computer Education Open	2	0.443	0.012	0.02
JMIR Medical Education	2	0.005	0.01	0.008
BMC Medical Education	2	0.114	0.011	0.012
Education Sciences	2	0.416	0.012	0.021
Teaching and Teacher Education	2	0.048	0.011	0.014
Künstliche Intelligenz	2	0.168	0.011	0.019
System	2	0.061	0.011	0.011

International Journal of Human-Computer Interaction	2	0.743	0.012	0.017
International Journal of STEM Education	2	0.258	0.011	0.02
Journal of Computer Assisted Learning	2	0.27	0.012	0.021
Journal of Educational Computing Research	2	0.183	0.011	0.018
Proceedings of the Frontiers in Education Conference	2	0.189	0.011	0.02
Behaviour & Information Technology	2	0.974	0.012	0.021
IEEE Transactions on Education	2	0.035	0.011	0.015
Structural Equation Modeling: A Multidisciplinary Journal	2	0.511	0.012	0.015
Nature Machine Intelligence	2	0.319	0.012	0.012
International Journal of Child-Computer Interaction	2	0.005	0.01	0.014
TechTrends	2	0.204	0.011	0.019
MIS Quarterly	3	118.656	0.018	0.01
International Journal of Information Management	3	86.354	0.018	0.009
Journal of Marketing Research	3	207.032	0.019	0.011
Journal of the Academy of Marketing Science	3	63.737	0.018	0.009
Technology in Society	3	83.671	0.017	0.01
Business Horizons	3	32.406	0.017	0.008

Fuente: Elaboración propia.

El Grupo 2 agrupa fuentes especializadas en educación y tecnología, como Computers and Education: Artificial Intelligence y Educ Inf Technol. Su relevancia estructural (alto PageRank) contrasta con su baja intermediación, indicando relevancia temática pero conexiones limitadas. Destaca Sustainability, que vincula IA con sostenibilidad, aunque la mayoría de los nodos (ej. JMIR Med Educ, Intermediación = 0.005) muestran aislamiento, dependiendo de pocas revistas centrales.

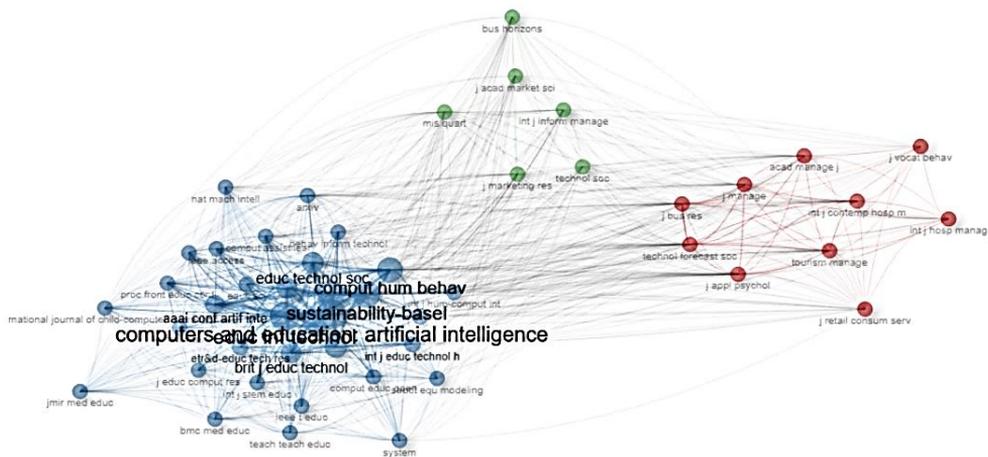
El Grupo 1 integra revistas de psicología aplicada y gestión, como Technol Forecast Soc y J Bus Res, que conectan teoría y aplicaciones empresariales.

Producción científica sobre alfabetización en inteligencia artificial

Su PageRank moderado sugiere influencia equilibrada, aunque nodos como Int J Hosp Manag revelan brechas en áreas como turismo o hostelería.

Figura 5

Mapa de cocitación de fuentes



Fuente: Elaboración propia.

Estructura social

La red de colaboración internacional, como se aprecia en la Tabla 16 y la Figura 6, se organiza en 12 grupos, con roles definidos por su capacidad de conexión y relevancia estructural. El Grupo 3 destaca como núcleo central, liderado por China, que actúa como principal conector global al integrar a países como Alemania, Reino Unido e Irán. Aunque su influencia es notable, la baja intermediación de naciones como Malasia evidencia desigualdades dentro del grupo. En paralelo, el Grupo 5, dominado por Estados Unidos, muestra fortaleza en vínculos transnacionales, pero la escasa intermediación de España o Bélgica sugiere dinámicas jerárquicas o colaboraciones limitadas. Los Grupos 2 y 4 representan conectores regionales. En el Grupo 2, Arabia Saudita e Italia funcionan como ejes en Oriente Medio, mientras Egipto y Kuwait tienen roles marginales. El Grupo 4, con Pakistán y Francia, emerge como puente entre Asia y Europa. Por otro lado, el Grupo 1, con Países Bajos como único conector relevante, refleja fragmentación en Europa y Asia meridional. Los Grupos 6 al 12 exhiben aislamiento estructural. Chile y Ecuador (Grupo 6), por ejemplo, tienen alta cohesión interna, pero sin conexión externa. Japón, Jordania o Azerbaiyán (Grupos 7-12) presentan métricas mínimas, indicando participación casi nula en la red global.

Esta estructura revela una jerarquía geopolítica clara: los Grupos 3 y 5 (China y EE.UU.) dominan con alta influencia; los Grupos 2 y 4 son conectores emergentes; y los Grupos 1, 6-12 representan periferia aislada (Sudamérica,

África, Asia Central). Las brechas son evidentes: regiones como África subsahariana están subrepresentadas, y economías avanzadas como España o Japón muestran baja integración. La predominancia de China y EE.UU. subraya su papel en moldear el campo, pero plantea desafíos para diversificar liderazgos y enfoques interculturales. Para abordar estas desigualdades, se requieren políticas que fomenten redes inclusivas y agendas globales más equitativas, integrando a países marginados y potenciando colaboraciones multisectoriales.

Tabla 16
Colaboración entre países

Nodo	Grupo	Intermediación	Cercanía	PageRank
Países Bajos	1	37	0.01	0.019
Bangladés	1	0	0.01	0.012
Arabia Saudita	2	46.223	0.013	0.035
Egipto	2	3.614	0.01	0.017
Austria	2	5.172	0.011	0.013
Italia	2	8.497	0.011	0.016
Kuwait	2	0	0.01	0.008
China	3	199.777	0.016	0.114
Corea	3	0.773	0.013	0.023
Alemania	3	53.49	0.015	0.051
Reino Unido	3	70.1	0.015	0.065
Canadá	3	26.398	0.013	0.025
Irán	3	41.271	0.014	0.034
Australia	3	18.957	0.014	0.041
Singapur	3	3.65	0.013	0.025
Malasia	3	0	0.012	0.021
Vietnam	3	3.158	0.013	0.021
Finlandia	3	13.947	0.013	0.028
India	3	6.554	0.013	0.022
Suecia	3	18.424	0.015	0.033
Noruega	3	2.146	0.013	0.021
Suiza	3	0.555	0.011	0.012
Catar	3	6.918	0.014	0.028
Ghana	3	0	0.011	0.008
Sudáfrica	3	0.884	0.011	0.013
Pakistán	4	118.286	0.014	0.036
Francia	4	46.495	0.014	0.028

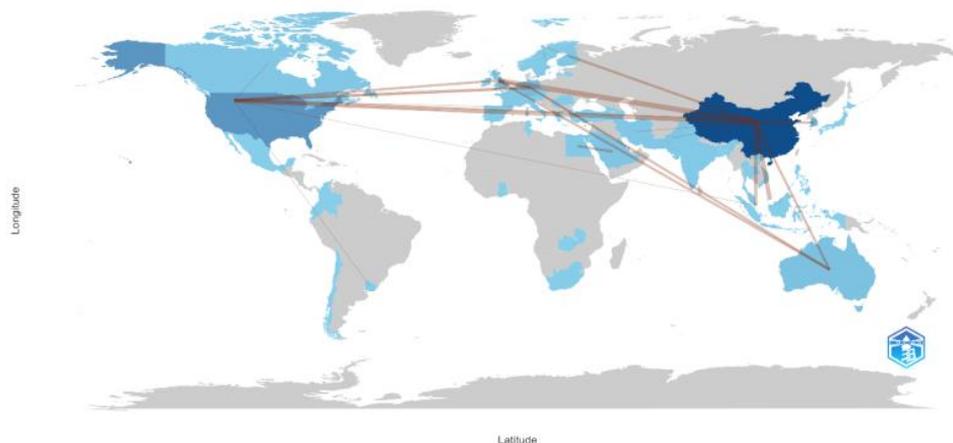
Producción científica sobre alfabetización en inteligencia artificial

Tailandia	4	3.573	0.013	0.016
EE.UU.	5	191.109	0.016	0.093
España	5	0	0.01	0.008
Bélgica	5	0.029	0.012	0.014
Uruguay	5	0	0.01	0.008
Indonesia	5	0	0.012	0.008
Chile	6	0	1	0.024
Ecuador	6	0	1	0.024
Israel	7	0	0.01	0.006
Azerbaiyán	8	0	0.009	0.006
Georgia	9	0	0.007	0.009
Irlanda	10	0	0.009	0.006
Japón	11	0	0.01	0.006
Jordania	12	0	0.01	0.006

Fuente: Elaboración propia.

Figura 6

Colaboración entre países
Country Collaboration Map



Fuente: Elaboración propia.

4. Discusión

El análisis de la producción científica sobre alfabetización en IA revela un campo de investigación en rápida expansión, especialmente evidente a partir de 2019 y con una aceleración notable en 2023-2024. Este crecimiento

exponencial coincide con la popularización de herramientas de IA y las crecientes demandas educativas en competencias digitales avanzadas (Benvenuti et al., 2023; Chen et al., 2020; Hashmi y Bal, 2024; Jackson y Jackson, 2024; Ng et al., 2023).

La concentración de un tercio de la producción científica en un núcleo reducido de 10 revistas, liderado por Education and Information Technologies, válida la Ley de Bradford (1934) y sugiere que estas fuentes son cruciales para acceder a la literatura clave.

El patrón de autoría, donde el 91,9% de los investigadores ha contribuido con un único documento mientras un pequeño núcleo impulsa la continuidad del área, es característico de campos emergentes. Autores como Chai CS, Chiu TKF y Su JH no solo son prolíficos, sino que también demuestran un impacto significativo y creciente en los últimos años.

Geográficamente, la investigación está liderada por China y Estados Unidos, que conjuntamente aglutinan más de la mitad de las publicaciones. Institucionalmente, un núcleo reducido de universidades, principalmente de Asia oriental, junto con entidades norteamericanas y europeas, encabeza la producción.

Temáticamente, la educación se erige como el área de investigación predominante y actúa como un puente interdisciplinario fundamental. Las palabras claves más frecuentes, inteligencia artificial y alfabetización en inteligencia artificial, confirman la centralidad de estos conceptos. Los análisis de redes de coocurrencia de palabras claves y áreas de investigación muestran una estructura con nodos centrales influyentes (como IA y educación) y otros más periféricos o especializados, lo que indica áreas de consolidación y otras emergentes o menos integradas. Las redes de cocitación y colaboración internacional refuerzan la posición dominante de ciertas revistas, países (China y EE. UU.) y la existencia de grupos de colaboración con diferentes grados de cohesión e influencia global. La evolución histórica de las citas de referencias también confirma la madurez reciente y la relevancia actual del campo.

Los hallazgos tienen implicaciones significativas. Para la educación, la centralidad de la investigación educativa y el enfoque en contextos como la educación K-12 y superior subrayan la necesidad de desarrollar e implementar currículos de alfabetización en IA adaptados a diferentes niveles. La identificación de temas clave como la ética en IA, la autoeficacia y la preparación para la IA debe guiar el diseño de contenidos formativos.

Para la investigación, el estudio identifica áreas temáticas consolidadas y otras más periféricas o emergentes, señalando oportunidades para futuras

Producción científica sobre alfabetización en inteligencia artificial

investigaciones que aborden nichos menos explorados o que busquen una mayor integración entre clústeres temáticos.

La predominancia de ciertos países e instituciones y las dinámicas de colaboración internacional sugieren la importancia de fomentar redes de investigación más diversificadas e inclusivas, integrando a países y regiones actualmente subrepresentados para enriquecer el campo con perspectivas interculturales. La identificación de autores y documentos influyentes puede ayudar a los investigadores a localizar trabajos fundamentales y potenciales colaboradores.

5. Conclusiones

Este estudio bibliométrico ha cartografiado el panorama de la investigación sobre alfabetización en IA en Web of Science hasta mayo de 2025, confirmando su estatus como un campo científico en vigorosa expansión, evidenciado por un aumento exponencial en la producción de artículos. El principal aporte de este trabajo radica en la provisión de una radiografía cuantitativa y estructural de este dominio emergente, identificando sus actores, focos temáticos y dinámicas de crecimiento e interconexión.

Por último, es fundamental reconocer las limitaciones inherentes a este estudio: su dependencia exclusiva de la base de datos Web of Science Core Collection, la exclusión de términos de búsqueda más amplios como machine learning, y la naturaleza instantánea del análisis en un campo de rápida evolución, además del limitado tiempo de citación para las obras más recientes.

Referencias

- Aria, Massimo y Cuccurullo, Corrado** (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. Disponible en: <https://doi.org/10.1016/j.joi.2017.08.007>
- Benvenuti, Martina, Cangelosi, Angelo, Weinberger, Armin, Mazzoni, Elvis, Benassi, Mariagrazia, Barbaresi, Mattia y Orsoni, Matteo** (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior*, 148, 107903. Disponible en: <https://doi.org/10.1016/j.chb.2023.107903>
- Bradford, Samuel Clement** (1934). Sources of information on specific subjects. *Engineering*, 137, 85-86.
- Brika, Said Khalfa, Algamdi, Abdelmageed, Chergui, Khalil, Musa, Adam y Zouaghi, Rabia** (2021). Quality of Higher Education: A Bibliometric Review Study. *Frontiers in Education*, 6. Disponible en: <https://doi.org/10.3389/feduc.2021.666087>

- Cetindamar, Dilek, Kitto, Kirsty, Wu, Mengjia, Zhang, Yi, Abedin, Babak y Knight, Simon** (2022). Explicating AI Literacy of Employees at Digital Workplaces. *IEEE Transactions on Engineering Management*, PP, 1-14. Disponible en: <https://doi.org/10.1109/TEM.2021.3138503>
- Chee, Hyunkyung, Ahn, Solmoe y Lee, Jihyun** (2024). A Competency Framework for AI Literacy: Variations by Different Learner Groups and an Implied Learning Pathway. *British Journal of Educational Technology*. Disponible en: <https://doi.org/10.1111/bjet.13556>
- Chen, Lijia, Chen, Pingping y Lin, Zhijian** (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264-75278. Disponible en: <https://doi.org/10.1109/ACCESS.2020.2988510>
- Chiu, Thomas K. F., Ahmad, Zubair, Ismailov, Murod y Sanusi, Ismaila Temitayo** (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*. Disponible en: <https://doi.org/10.1016/j.caeo.2024.100171>
- Clarivate Analytics** (s.f.). Web of Science Core Collection. Web of Science. Disponible en: <https://webofscience.help.clarivate.com/Content/wos-core-collection/wos-core-collection.htm>
- Donthu, Naveen, Kumar, Satish, Mukherjee, Debmalya, Pandey, Nitesh y Lim, Weng Marc** (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. Disponible en: <https://doi.org/10.1016/J.JBUSRES.2021.04.070>
- Ellegaard, Ole y Wallin, Johan** (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, 105, 1809-1831. Disponible en: <https://doi.org/10.1007/s11192-015-1645-z>
- Hashmi, Nada y Bal, Anjali** (2024). Generative AI in higher education and beyond. *Business Horizons*. Disponible en: <https://doi.org/10.1016/j.bushor.2024.05.005>
- Jackson, Emerson y Jackson, Hudson** (2024). Enhancing Human Knowledge and Capabilities with Artificial Intelligence Tools for Education. *Educational Challenges*. Disponible en: <https://doi.org/10.34142/2709-7986.2024.29.2.09>
- Jantakun, Kitsadaporn, Jantakun, Thiti y Jantakoon, Thada** (2025). Bibliometric Analysis of Artificial Intelligence for Digital Literacy. *Journal of Education and Learning*, 14(3), 115. Disponible en: <https://doi.org/10.5539/jel.v14n3p115>
- Jing, Yuhui, Wang, Chengliang, Chen, Yu, Wang, Haoming, Yu, Teng y Shadiev, Rustam** (2023). Bibliometric mapping techniques in educational technology research: A systematic literature review. *Education and Information Technologies*, 29, 9283-9311. Disponible en: <https://doi.org/10.1007/s10639-023-12178-6>
- Khan, Ashraf, Goodell, John Hassan, Kabir y Paltrinieri, Andrea** (2021). A bibliometric review of finance bibliometric papers. *Finance Research Letters*. Disponible en: <https://doi.org/10.1016/j.frl.2021.102520>

Producción científica sobre alfabetización en inteligencia artificial

- Laupichler, Matthias, Aster, Alexandra, Schirch, Jana y Raupach, Tobias** (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101. Disponible en: <https://doi.org/10.1016/j.caeai.2022.100101>
- Lotka, Alfred** (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317-323.
- Mukherjee, Debmalaya, Lim, Weng Marc, Kumar, Satish y Donthu, Naveen** (2022). Guidelines for advancing theory and practice through bibliometric research. *Journal of Business Research*. Disponible en: <https://doi.org/10.1016/j.jbusres.2022.04.042>
- Ng, Davy Tsz Kit, Leung, Jac Ka Lok, Su, Jiahong, Ng, Ross Chi Wui y Chu y Samuel Kai Wah** (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, 71, 137-161. Disponible en: <https://doi.org/10.1007/s11423-023-10203-6>
- Passas, Ioannis** (2024). *Bibliometric Analysis: The Main Steps*. Encyclopedia. Disponible en: <https://doi.org/10.3390/encyclopedia4020065>
- Pinski, Marc y Benlian, Alexander** (2024). AI literacy for users – A comprehensive review and future research directions of learning methods, components, and effects. *Computers in Human Behavior: Artificial Humans*. Disponible en: <https://doi.org/10.1016/j.chbah.2024.100062>
- Salinas-Ríos, Karla y López, Angélica** (2022). Bibliometrics, a useful tool within the field of research. *Journal of Basic and Applied Psychology Research*. Disponible en: <https://doi.org/10.29057/jbapr.v3i6.6829>
- Son, Mihyun y Jho, Hunkoog** (2024). Analyzing Research Trends and Relationships in AI Literacy, Digital Literacy, Data Literacy, and Knowledge Information Processing Competency through Bibliometric Analysis. *Brain, Digital, & Learning*. Disponible en: <https://doi.org/10.31216/bdl.20240022>
- Sperling, Katarina, Stenberg, Carl-Johan, McGrath, Carmac, Åkerfeldt, Anna, Heintz, Fredrik y Stenliden, Linnéa** (2024). In search of artificial intelligence (AI) literacy in Teacher Education: A scoping review. *Computers and Education Open*. Disponible en: <https://doi.org/10.1016/j.caeo.2024.100169>
- Stolpe, Karin y Hallström, Jonas** (2024). Artificial Intelligence Literacy for Technology Education. *Computers and Education Open*. Disponible en: <https://doi.org/10.1016/j.caeo.2024.100159>
- Wang, Bingcheng, Rau, Pei-Luen Patrick y Yuan, Tianyi** (2022). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42, 1324-1337. Disponible en: <https://doi.org/10.1080/0144929X.2022.2072768>

- Wu, Changhong, Zou, Jiatong, Wang, Shuyi, Sun, Bo y Zhang, Jingpeng** (2024). Bibliometric Study and Visualization Analysis of Domestic and International Information Literacy Research in the Era of Artificial Intelligence. *International Journal of Librarianship*. Disponible en: <https://doi.org/10.23974/ijol.2024.vol9.3.408>
- Yim, Iris y Heung Yim** (2024). A Critical Review of Teaching and Learning Artificial Intelligence (AI) Literacy: Developing an Intelligence-based AI Literacy Framework for Primary School Education. *Computers and Education: Artificial Intelligence*. Disponible en: <https://doi.org/10.1016/j.caeai.2024.100319>
- Zupic, Ivan y Čater, Tomaž** (2014). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, 18, 429-472. Disponible en: <https://doi.org/10.1177/1094428114562629>

Narrativas Distópicas y Utópicas: Uso de la IA en la Arquitectura.

Dystopian and Utopian Narratives: Using AI in Architecture.

Debby Avendaño Sánchez
debbyavendano@gmail.com

Resumen

El artículo analiza las narrativas contrapuestas sobre el uso de la inteligencia artificial (IA) en el campo de la arquitectura, abordando su potencial transformador como herramienta aliada del diseño y, a su vez, sus implicancias éticas, sociales y ontológicas cuando es percibida como una amenaza. Se utiliza un enfoque crítico-discursivo apoyado en el análisis cultural e histórico de las narrativas sobre tecnología y arquitectura. Se desarrolla una revisión documental detallada sobre los aspectos que nos ocupan, se examinan casos emblemáticos, conceptos filosóficos y producciones mediáticas que alimentan el imaginario utópico y distópico de la IA. Se estructura el estudio en dos grandes ejes narrativos: las afirmaciones que celebran la integración tecnológica y las críticas que advierten sobre su uso deshumanizante. Los resultados revelan tensiones entre eficiencia algorítmica y sensibilidad proyectual, junto con riesgos vinculados a la deshumanización del trabajo, la estética simulada, la precarización profesional y la pérdida de sentido disciplinar. También se reconocen aportes transformadores de la IA en términos de visualización, diseño asistido y generación de propuestas creativas. Se destaca, además, que las narrativas distópicas no solo critican el impacto de la IA, sino que históricamente han contribuido a cuestionar y renovar el pensamiento arquitectónico.

Palabras claves: arquitectura, inteligencia artificial, narrativa, distopía, utopía.

Abstract

This article analyzes the conflicting narratives about the use of artificial intelligence (AI) in the field of architecture, addressing its transformative potential as an allied tool for design and, in turn, its

ethical, social, and ontological implications when it is perceived as a threat. A critical-discursive approach is used, supported by a cultural and historical analysis of narratives about technology and architecture. A detailed documentary review is conducted on the issues at hand, examining emblematic cases, philosophical concepts, and media productions that fuel the utopian and dystopian imaginary of AI. The study is structured around two main narrative axes: statements that celebrate technological integration and critiques that warn against its dehumanizing use. The results reveal tensions between algorithmic efficiency and design sensitivity, along with risks linked to the dehumanization of work, simulated aesthetics, professional precarization, and the loss of disciplinary meaning. The transformative contributions of AI are also recognized in terms of visualization, assisted design, and the generation of creative proposals. It is also highlighted that dystopian narratives not only critique the impact of AI but have historically contributed to questioning and renewing architectural thinking.

Keywords: architecture, artificial intelligence, narrative, dystopia, utopia.

1. Introducción

Lo inmediato es confuso y la historia necesita perspectiva temporal. Los acontecimientos del tema que aquí nos atañe, están en pleno desarrollo. Solo en el transcurso de los años se podrán evaluar las consecuencias, continuidades y cambios estructurales de los procesos desarrollados a partir de la implementación de la Inteligencia Artificial (IA) en la Arquitectura.

La creación, implementación e influencia de las tecnologías disruptivas acusan recibo de un fenómeno comunicacional superficial y contradictorio saturado de versiones, opiniones, desinformación y propaganda, distintivo de la inmediatez en la era digital; paralelamente ha desatado también un creciente debate académico que aspira dilucidar sus alcances y límites, así como su impacto negativo o positivo.

En este artículo se busca una aproximación a estos escenarios narrativos que por un lado encuentran en la IA una aliada, y por el otro, una amenaza. Las narrativas desarrolladas a consecuencia de esta dualidad, manejan tanto escenarios con elementos distópicos que cuestionan el presente como utópico que proyectan un mundo ideal.

A pesar de estas visiones contrastantes y sus polaridades intrínsecas, destaca que ambas esgrimen con certeza una idea común: la integración de

Debby Avendaño Sánchez

la inteligencia artificial en la arquitectura está transformando la formación académica y el ejercicio de la profesión.

¿Por qué resulta de interés conocer o contraponer esta doble narrativa?

Porque demuestra que el conocimiento, el desarrollo y uso de tecnologías, así como los discursos que desatan en consecuencia están atravesados por intereses, prejuicios (ideológicos), y contextos.

La historia ha expuesto que las revoluciones industriales y científicas no sólo transformaron el mundo material de la sociedad, sino que también generaron **narrativas contrapuestas** que reflejaron conveniencias, ideologías y visiones del progreso muy distintas.

Ejemplos pertinentes para ilustrar esta dicotomía se observan tanto en la Revolución científica del siglo XVII como la Industrial del siglo XVIII, cuyas narrativas anexas manejaban un sentido positivo sobre el progreso o una visión crítica del mismo.

En el primer lugar, hay una perspectiva optimista y racionalista, liberadora del pensamiento que promueve el desarrollo, la modernización y celebra las innovaciones; en el segundo, cobra valor la óptica de la resistencia que manifiesta tanto conflictos planteados por instituciones o sus representantes de tendencia conservadora, o el rechazo a condiciones adversas (de índole social, ecológico, entre otros), producto de esas mismas innovaciones.

La aproximación a estas narrativas históricas confrontadas permite entender cómo se legitimaron ciertos discursos, y además, proporcionan una mirada más crítica y plural sobre estos fenómenos o procesos del pasado.

Con la ambición de esa misma mirada, se aborda el caso que atañe a este escrito, consciente de los límites que dicha empresa conlleva, debido a que constituye un proceso en pleno desarrollo con transformaciones e innovaciones que se suceden unas a otras, ampliando sus facultades en breves periodos de tiempo. Al igual que en las revoluciones industriales y científicas, la IA genera visiones enfrentadas sobre su impacto en la vida. En la actualidad, su uso en la sociedad está rodeado de **narrativas contrapuestas** que reflejan tanto entusiasmo como preocupación, dupla que aplica también para el ámbito narrativo de la arquitectura.

2. La IA como aliada: del diseño asistido al imaginario idealista

„En el umbral de una nueva era, la arquitectura se encuentra en un punto de inflexión que no sólo redefine su práctica, sino que también rediseña su

Narrativas Distópicas y Utópicas: Uso de la IA en la Arquitectura

futuro“, como afirma Rossel (2024, 4). Esta transformación, impulsada por el auge de la inteligencia artificial (IA), se despliega en un escenario marcado por cierta incertidumbre epistemológica, pero también por un abanico de narrativas optimistas que destacan sus aportes estratégicos y operativos dentro de la disciplina.

La promoción de la IA como aliada fundamental en el ámbito arquitectónico se sustenta, ante todo, en sus atributos funcionales. Las nuevas herramientas digitales permiten optimizar procesos que abarcan desde la concepción inicial y la planificación detallada hasta la ejecución constructiva. Su capacidad para procesar grandes volúmenes de datos, reconocer patrones complejos y generar alternativas de diseño en tiempo récord resulta particularmente atractiva frente a las dinámicas vertiginosas que caracterizan al mercado laboral y al sector inmobiliario contemporáneo, como se manifiesta desde las plataformas Hypar (2022) y TestFit (2022).

Esta irrupción tecnológica, al reducir significativamente la inversión temporal requerida en etapas claves del proyecto, está produciendo una transformación profunda en el ejercicio profesional. No se trata únicamente de incorporar nuevos recursos operativos, sino de reformular las propias competencias del arquitecto, integrando habilidades creativas, analíticas y colaborativas dentro de un marco interdisciplinario altamente digitalizado, tal como explican Roudsari y Pak (2013).

Tal transformación es asumida por las narrativas positivas no como una ruptura, sino como una evolución del perfil profesional. Estas perspectivas defienden la tesis de que la IA aporta beneficios sustanciales al quehacer arquitectónico, posicionándola no sólo como una herramienta funcional del presente, sino como un agente de cambio con proyecciones hacia el futuro.

Destaca Valencia (2023) que, lejos de representar una amenaza, la inteligencia artificial puede convertirse en un aliado estratégico que expande las fronteras del diseño, fomenta la eficiencia y promueve enfoques creativos antes impensables. Este enfoque visionario, por momentos idealista, imagina un rol renovado para el arquitecto, pero también corre el riesgo de deslizarse hacia modelos utópicos, donde la tecnología parece resolver de forma automática las complejidades proyectuales, sociales y ambientales.

A continuación, se examinarán los principales aportes de la inteligencia artificial en arquitectura según estas narrativas afirmativas, incorporando ejemplos aplicados que evidencian su incidencia en distintas fases del proceso proyectual. Uno de los aportes más significativos de la IA es la expansión de la creatividad arquitectónica. Las herramientas generativas basadas en redes neuronales permiten explorar morfologías y configuraciones espaciales que difícilmente surgirían mediante métodos tradicionales. Estas tecnologías no solo amplían el repertorio formal del

Debby Avendaño Sánchez

arquitecto, sino que también estimulan procesos de diseño más especulativos e innovadores. Plataformas como *Midjourney* o *DALL·E* generan visualizaciones arquitectónicas a partir de descripciones textuales, ofreciendo un abanico de posibilidades estéticas que pueden servir como punto de partida conceptual (OpenAI, 2023).

Figura 1

La imagen puede simbolizar el proceso de creación que caracteriza la IA generativa



Nota: Patrones repetitivos y estructura en capas. **Fuente:** Microsoft 365 Stock Images (2025).

Además, la IA actúa como una „mente colaborativa“, concepto que remite a la idea de inteligencia compartida entre humanos y sistemas digitales. Esta noción, vinculada a la inteligencia colaborativa propuesta por Pierre Lévy (1994), sugiere que el arquitecto no trabaja en solitario, sino que dialoga con algoritmos capaces de analizar variables complejas, prever escenarios y sugerir soluciones. Herramientas como *Hypar* o *TestFit* permiten simular configuraciones urbanas en segundos, considerando parámetros como densidad, accesibilidad y eficiencia en movilidad. Así, el arquitecto puede enfocarse en decisiones estratégicas, mientras delega tareas analíticas a sistemas inteligentes, como lo exponen Choudhury y Chowdhury (2023).

La incorporación de IA también exige el desarrollo de nuevas competencias profesionales. El arquitecto contemporáneo debe familiarizarse con lenguajes de programación, análisis de datos y modelos de *machine learning*, lo que representa una evolución del perfil tradicional hacia una práctica interdisciplinaria. El uso de plataformas como *Grasshopper*, junto con plugins como *Ladybug Tools*, facilita la simulación sobre el comportamiento ambiental de las futuras edificaciones, integrando datos climáticos y energéticos en el

Narrativas Distópicas y Utópicas: Uso de la IA en la Arquitectura

proceso de diseño como señalan Roudsari y Pak (2013, 3132). Esta confluencia entre arquitectura, sostenibilidad y ciencia de datos, abre posibilidades novedosas no solo en la práctica profesional sino también en los campos de la investigación y formación académica.

En esta línea, la inteligencia artificial también impulsa una sostenibilidad inteligente, al permitir el análisis avanzado de entornos urbanos y naturales para generar propuestas más eficientes y respetuosas con el medio ambiente, basadas en una toma de decisiones más informada. Los algoritmos especializados son capaces de evaluar variables complejas como la topografía, el flujo vehicular, la orientación solar, la calidad del aire, la ventilación cruzada y el impacto acústico, optimizando las condiciones de confort ambiental desde las etapas iniciales del diseño.

En el 2021, la corporación Autodesk ejemplificaba este enfoque con la extinta plataforma *Spacemaker AI* que facilitaba la conceptualización de configuraciones residenciales maximizando la luz natural y reduciendo el consumo energético, en consonancia con los objetivos de desarrollo sostenible. Luego, su sucesora, *Autodesk Forma*, amplía significativamente estas capacidades al abordar el ciclo completo de vida del proyecto, desde su concepción hasta la construcción. Además, permite detectar anticipadamente fallos de diseño antes de la ejecución, y se integra eficazmente con otras plataformas BIM¹ como *Revit*, reforzando su valor estratégico dentro del enfoque sostenible de la arquitectura contemporánea, como lo afirma Autodesk (2025).

Por otro lado, la automatización de tareas técnicas mediante IA libera al arquitecto de procesos repetitivos como la elaboración de planos, cálculos estructurales o modelado 3D. Herramientas como *Revit* incorporan diseño generativo que permite automatizar la disposición de elementos arquitectónicos a partir de parámetros definidos, mejorando la eficiencia sin comprometer la calidad del diseño, según Autodesk (2020). Esta automatización no implica una pérdida de control, sino una redistribución del esfuerzo hacia aspectos más estratégicos y creativos.

Finalmente, la IA permite optimizar el tiempo de diseño al generar múltiples propuestas en cuestión de minutos. Los modelos de *machine learning* pueden aprender del historial de proyectos anteriores y proponer configuraciones que han demostrado ser funcionales y estéticas. Esto acelera la toma de decisiones y mejora la productividad en estudios con alta demanda, sin

¹ **Nota de la Autora:** BIM: Building Information Modeling, por sus siglas en inglés. Este software de Modelado de Información para la Construcción comprende un proceso que integra a todos los involucrados en un proyecto arquitectónico y su ejecución. El método de trabajo es colaborativo compartiendo información en un modelo digital.

Debby Avendaño Sánchez

sacrificar la calidad ni la originalidad del diseño, como lo promueve la plataforma TestFit (2022).

En resumen, las narrativas que se apoyan exclusivamente en los aportes de la inteligencia artificial en arquitectura sostienen, en clave afirmativa, que esta tecnología no solo transforma las herramientas operativas del arquitecto, sino también su modo de pensar, colaborar y proyectar.

Estas ópticas favorecedoras visualizan una arquitectura del futuro necesariamente híbrida, donde la inteligencia humana y artificial convergen en un diálogo creativo, ético y sostenible. No obstante, dicho horizonte proyectual no está exento de matices idealistas, ya que se tiende a imaginar un escenario donde la IA actúa como catalizador de soluciones perfectas, capaces de resolver con precisión algorítmica los desafíos del diseño, la planificación urbana y la sostenibilidad ambiental.

En este marco, la IA es exaltada como una oportunidad transformadora, una herramienta potencial al servicio del arquitecto y disciplinas afines, que impulsa un salto evolutivo más que una ruptura revolucionaria. Esta visión, de tintes utópicos, en ocasiones atribuye a la figura del arquitecto un rol casi mesiánico (como ha sucedido en otras épocas de la historia), agente visionario capaz de liderar el cambio y reconfigurar el entorno construido.

Así lo interpreta Rossel (2024), al citar a Gamage, Hyde y Newton (2021, 215): „Aquellos que logren adaptarse y liderar el cambio serán los arquitectos del futuro, visionarios capaces de transformar las ciudades y los espacios de acuerdo a las necesidades de una sociedad en rápida transformación“. Esta afirmación refuerza la idea de que el arquitecto del mañana no solo debe dominar las herramientas digitales, sino también asumir una postura anticipatoria frente a los desafíos que plantea la tecnificación del hábitat.

3. Creación del imaginario distópico: IA entre la deshumanización y el simulacro

En las últimas décadas, nuevas narrativas especulativas, distópicas y (post)apocalípticas han tomado fuerza en el discurso cultural global, y el impacto transformador de los productos multi/transmedia en la sociedad parece haberlas potenciado. Estas narrativas han salido del campo restringido de la literatura y el cine para habitar diversas formas artísticas y mediáticas, alcanzando una resonancia global creciente y una difusión sin precedentes. La arquitectura como producto cultural acusa recibo de este discurso envolvente porque el hábitat —y la manera como se concibe y diseña— responde a los modos de vida.

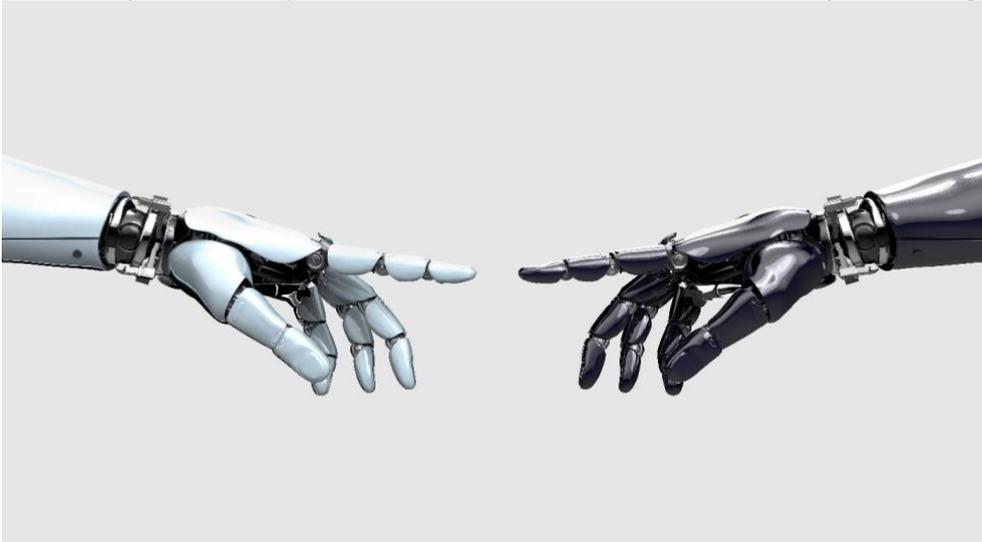
Narrativas Distópicas y Utópicas: Uso de la IA en la Arquitectura

El interés en narrativas que incorporan elementos distópicos como crítica o advertencia sobre el uso de la IA en la arquitectura viene a contrastar la perspectiva idealizada registrada en el apartado anterior. El imaginario arquitectónico ha contado ya con escenarios urbanos distópicos presentados en películas como „Metropolis“ (1927), o „Blade Runner“ (1982), las cuales reflejan preocupaciones de índole social, político y ecológico, en medio de megaciudades altamente tecnificadas.

La distopía permite cuestionar el presente y advierte sobre escenarios futuros si no se revisa críticamente la realidad. En la narrativa arquitectónica contemporánea, este enfoque se traduce en reflexionar sobre cómo se habita, cómo se diseña y cuáles valores se priorizan tras la disrupción que conlleva la aplicación de la IA en el diseño arquitectónico; desde allí se plantean interrogantes no sólo sobre la eficiencia y la optimización de procesos sino especialmente sobre el rol y el valor de la creatividad humana frente a la lógica algorítmica, así como su impacto tanto en los procesos cognitivos del arquitecto-creador cómo en la experiencia humana del habitar.

Figura 2

El contraste y la dualidad presentes en la relación entre el ser humano y la tecnología



Nota: Interacción entre dos manos robóticas. **Fuente:** Microsoft 365 Stock Images (2025).

Esta perspectiva que concibe a la inteligencia artificial como una amenaza se inscribe en una línea discursiva que le antecede en el ámbito arquitectónico. Diversas corrientes críticas han interpretado la arquitectura contemporánea como resultado de una „inversión de los ideales utópicos“ promovidos por el „Movimiento Moderno“.

Debby Avendaño Sánchez

Aquellos elementos que en las vanguardias de los años 1920 y las neovanguardias de los años 1960 eran exaltados como emblemas de progreso —las megaciudades, la automatización y la eficiencia sistematizada— son actualmente reinterpretados como dispositivos de control espacial, generadores de alienación y propiciadores de la erosión identitaria como señalan Meléndez (2023) y López (1991).

Desde esta perspectiva crítica, se habilita un marco narrativo que interpela las implicancias éticas, culturales y ontológicas del quehacer arquitectónico en el contexto de la revolución tecnológica impulsada por la IA. Gamage, Hyde y Newton (2021, 216) advierten que, en ausencia de una gestión mediada por la conciencia crítica y el pensamiento reflexivo, la tecnología podría devenir en amenaza, profundizando dilemas que trascienden lo técnico para instalarse en el centro de los debates éticos contemporáneos. A continuación, se presentan algunos elementos fundamentales que contribuyen a desarrollar este enfoque:

Deshumanización del trabajo: si la IA reemplaza progresivamente las funciones cognitivas humanas, puede afectar la identidad laboral, la sinergia del trabajo en equipo y el sentido de utilidad, generando un desarraigo emocional, como advierte Rossel (2024, 4871).

Condicionamiento o pérdida de creatividad en el proceso de diseño: a medida que los algoritmos generan propuestas basadas en parámetros predefinidos y patrones de datos, se plantea la preocupación de que los resultados sean estandarizados y carentes de singularidad. Paradójicamente, la IA también puede sugerir ideas insospechadas y soluciones excepcionales; sin embargo, existe el riesgo de que el diseñador desarrolle una dependencia excesiva de sus propuestas, tal como advierten Choudhury y Chowdhury (2023, 106).

Sesgos algorítmicos: los sistemas de IA pueden reproducir prejuicios sociales, estéticos o culturales, al aprender de datos históricos sesgados, lo que podría derivar en modelos excluyentes, como se advierte desde la plataforma de OpenAI (2023, 11).

Disminución del valor del trabajo de los diseñadores: según Mamani (2023), la automatización podría desplazar profesionales cuya labor puede ser reemplazada por algoritmos generativos que ofrecen una amplia gama de opciones de diseño en menos tiempo.

Narrativas engañosas: al exagerar las capacidades de la IA, se pueden generar vínculos ilusorios con la tecnología, minimizando el trabajo humano detrás de su desarrollo, como señala Valencia (2023).

Hipercomunicación visual: la visualización de proyectos se ha intensificado vertiginosamente con la integración de la IA en tecnologías como la realidad aumentada. Arquitectos y clientes interactúan en tiempo real con modelos tridimensionales de los proyectos, tomando decisiones y realizando ajustes instantáneos, como muestran Hypar (2022) y Autodesk (2025).

Uso de la tecnología sin ética: el empleo de software paramétrico y algoritmos que priorizan la forma sobre la habitabilidad puede derivar en espacios deshumanizados, desvinculados de la sociedad y su entorno. Esta preocupación ha sido documentada en los análisis sobre estética computacional y diseño especulativo, como señalan Roudsari y Pak (2013, 3132).

El edificio inteligente como dispositivo de control: las edificaciones que recopilan datos de sus usuarios —temperatura, movimiento, prácticas, usos— pueden mejorar la eficiencia y la seguridad, pero también invadir la privacidad si no se establecen protocolos éticos claros, como advierten Lévy (1994), y la plataforma Autodesk (2021).

Desigualdad y precarización: el desconocimiento sobre la IA podría profundizar la vulnerabilidad social y laboral, especialmente en contextos donde se consume pero no se produce tecnología, como advierte Mamani (2023).

Estos riesgos asociados a la IA en el campo de la arquitectura, como los denomina Mamani (2023), han tenido sus correlatos históricos en cada época de cambios significativos y de avances tecnológicos. La narrativa crítica y de resistencia ha tenido su espacio como bastión de defensa de la misma disciplina y del habitar humano. Esto resulta importante porque la arquitectura no es solo un producto de la materia y la técnica, sino también de la cultura, la política y la ética, tal como sostiene Frampton (1995, 19), al resaltar que toda producción arquitectónica responde a valores culturales y contextuales, y no puede desligarse de sus condicionamientos históricos. Cada decisión tomada (así esté sugerida por una plataforma tecnológica), desde el diseño hasta su construcción, tiene consecuencias directas en el modo en que se vive y se sienten los espacios.

Todos estos riesgos pueden tomarse como desafíos, lo que convoca a la necesidad de repensar las competencias que requieren los profesionales del sector, tanto los que están ejerciendo como los que están en formación. En este sentido, Gamage, Hyde y Newton (2021, 218) plantean que los arquitectos deben reimaginar su identidad profesional (repensar las nuevas competencias que requieren), en un entorno digital dinámico, redefiniendo no solo sus herramientas, sino también la naturaleza de su trabajo y el alcance ético de sus decisiones.

Este recorrido por algunos elementos de las narrativas distópicas aplicadas al campo arquitectónico revela una dimensión crítica que se opone —con fundamento ético y cultural— a la visión optimista y tecnocéntrica que tiende a dominar los discursos contemporáneos sobre inteligencia artificial. Estas perspectivas, más que ser profecías negativas, operan como advertencias poderosas que invitan a cuestionar la dirección en que avanza la disciplina, alertando sobre los riesgos de una práctica despojada de humanidad, singularidad y sentido social. Frente a los escenarios de alienación, estandarización y vigilancia que podrían instalarse si no se asumen posturas conscientes, emerge la necesidad urgente de construir una arquitectura más reflexiva, donde la creatividad humana, la responsabilidad ética y el compromiso con el habitar sean valores irrenunciables. Esta narrativa distópica no busca cancelar la innovación, sino reinscribirla en un horizonte crítico donde imaginar futuros posibles incluya siempre el derecho a habitar con dignidad.

4. Conclusiones

Cómo síntesis de lo expuesto, al recuperar las voces contrapuestas que atraviesan este debate, tanto las narrativas afirmativas como las distópicas permiten pensar la arquitectura contemporánea como un campo tensionado entre promesas de innovación y advertencias de deshumanización. Vivimos rodeados de interrogantes, y la inteligencia artificial —como toda tecnología disruptiva— se inserta en ese tejido de incertidumbre con fuerza transformadora. Su aplicación en el ámbito arquitectónico refleja esa dualidad: puede expandir el proceso creativo, facilitar decisiones complejas y enriquecer la experiencia proyectual; pero también puede devenir en amenaza si no se gestiona con conciencia crítica, poniendo en riesgo la singularidad, la sensibilidad y el vínculo humano con el espacio.

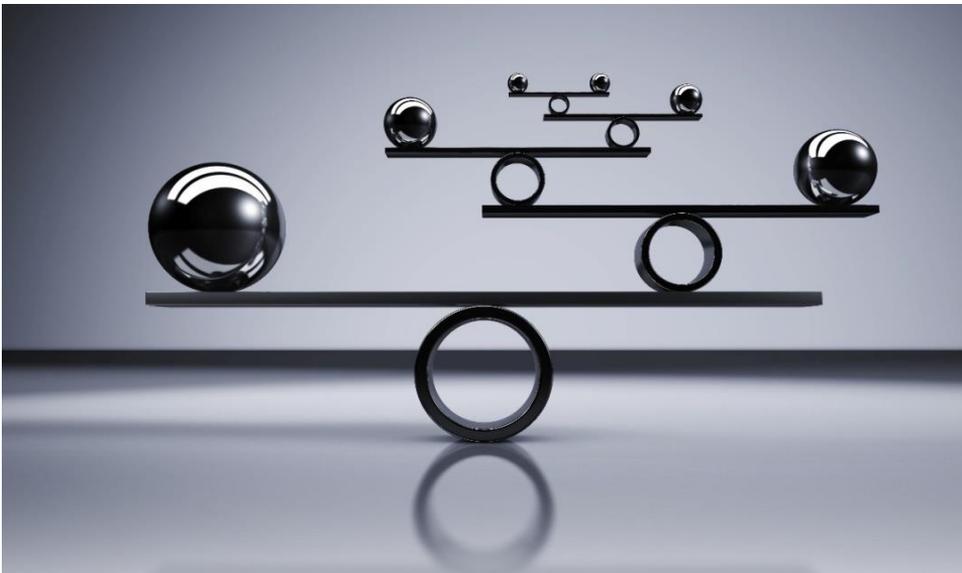
Estas narrativas importan porque visibilizan los dilemas éticos, sociales y ontológicos que plantea la integración de la IA, y porque invitan a construir una mirada más plural y reflexiva sobre sus usos. También interpelan la necesidad de educar, regular y posicionar la práctica arquitectónica en defensa de su dimensión ética, cultural y emocional. La arquitectura no es un conjunto de imágenes espectaculares para redes ni simulacros renderizados de habitabilidad aparente —como advertía Baudrillard—; es, ante todo, una forma de conectar con las personas y enriquecer sus vidas. La estética como simulacro, la desmaterialización del espacio y la pérdida del aura benjaminiana representan riesgos tangibles cuando se prioriza la representación por encima de la experiencia, generando edificios genéricos, espectaculares pero desvinculados de su contexto.

Frente a esta encrucijada, el arquitecto contemporáneo debe replantearse su rol, no solo como operador de herramientas digitales, sino como garante de

sentido, como mediador entre algoritmos y habitabilidad. La Cuarta Revolución Industrial —como señala Schwab— enfrenta posturas polarizadas, pero en este campo, como en muchos otros, la respuesta no reside en los extremos. El futuro de la arquitectura dependerá de quién diseña, para quién y con qué valores. Quizá el verdadero progreso no consiste en automatizar el oficio, sino en imaginar nuevas formas de creatividad compartida entre humanidad y máquina, en las que el espacio arquitectónico conserve su capacidad de ser vivido, sentido e incluso habitado antes de ser construido.

Figura 3

Disposición asimétrica en equilibrio visual y físico. La respuesta no reside en un solo extremo



Nota: Escultura en equilibrio. Abstracción geométrica. **Fuente:** Microsoft 365 Stock Images (2025).

Referencias

- Autodesk** (2021). Spacemaker AI: Urban design powered by artificial intelligence. Disponible en: <https://www.spacemaker.ai>
- Autodesk** (2025). Autodesk Forma: Diseño conceptual y análisis ambiental impulsado por IA. Disponible en: <https://www.autodesk.com/es/products/forma/overview>
- Choudhury, Anirban y Chowdhury, Rittika** (2023). „Digital transformation in architecture: Impacts of AI-driven design tools“. En: International Journal of Architectural Technology, 8(2), 98–112.
- Datech** (2023). Autodesk y la Inteligencia Artificial en Arquitectura. Disponible en: <https://www.datech.es/software/autodesk-e-inteligencia-artificial-en-arquitectura>

- Frampton, Kenneth** (2024). Historia crítica de la arquitectura moderna. Quinta edición revisada y ampliada. Barcelona, España: Editorial GG.
- Gamage, Dinusha, Hyde, Rory y Newton, Paul** (2021). „Architects of the future: Reimagining professional identity in the age of digital transformation“. En: Journal of Architectural Education, 75(2), 215–229. Disponible en: <https://doi.org/10.1080/10464883.2021.1901234>
- Hypar** (2022). Hypar Platform Overview. Disponible en: <https://www.hypar.io>
- Lévy, Pierre** (1994). L'intelligence collective: Pour une anthropologie du cyberspace. París, Francia: La Découverte.
- Mamani, Gustavo** (2023). „El impacto negativo del uso de la inteligencia artificial en el proceso de diseño arquitectónico“. En: Revista de Arquitectura y Urbanismo Taypi, Vol. 2, N° 1. Puno, Perú, pp. 10–12.
- OpenAI** (2023). *GPT-4 Technical Report*. Disponible en: <https://openai.com/index/gpt-4-research>
- Rossel, Marcelo** (2024). „La arquitectura en el umbral de la inteligencia artificial: Narrativas, transformaciones y desafíos“. En: LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, 5(5), 4869–4883. Disponible en: <https://doi.org/10.56712/latam.v5i5.2949>
- Roudsari, Mostapha y Pak, Michelle** (2013). „Ladybug: A parametric environmental plugin for Grasshopper to help designers create an environmentally-conscious design“. En: Proceedings of Building Simulation 2013: 13th Conference of IBPSA, pp. 3128–3135. Recuperado de: https://publications.ibpsa.org/conference/paper/?id=bs2013_2499
- Schwab, Klaus** (2016). La Cuarta Revolución Industrial. Madrid, España: Editorial DEBATE.
- TestFit** (2022). Real Estate Feasibility Platform. Disponible en: <https://www.testfit.io>
- Valencia, Nicolás** (2023). El impacto de la inteligencia artificial en la arquitectura en 2024 (y más allá). Santiago, Chile: ArchDaily. Disponible en: <https://www.archdaily.cl/cl/1010103/el-impacto-de-la-inteligencia-artificial-en-la-arquitectura-en-2024-y-mas-alla>

Inteligencia Artificial: Análisis de su impacto en la brecha digital.

Artificial Intelligence: Analysis of its impact on the digital divide.

Beatriz Chuquimia Aranda
mscbjca@gmail.com

Resumen

Desde la irrupción del Internet y la comunicación a distancia mediante redes de computadoras, ha permitido que la información haya tenido una revolución importante en el medio informático. El acceso a los dispositivos electrónicos y al Internet en los años noventa, dieron lugar a la denominada brecha digital. Posteriormente estas brechas dieron lugar a diversos debates sobre las diferencias y desigualdades en entornos digitales, que progresivamente ya no se trata solamente de una división en el acceso, sino entre las condiciones de uso, habilidades y una serie de fracturas entre las que se encuentra la brecha digital en razón de territorio, por cuestiones económicas e incluso de género por mencionar algunas. En este escrito, se analiza el impacto que tiene la inteligencia artificial en la brecha digital, tomando en cuenta los aspectos socio-demográficos principales en relación de acceso a la banda ancha y al índice de precios en Bolivia y América Latina. Asimismo, para hallar resultados se ha efectuado ejercicios con dos plataformas de modelos de lenguaje con Inteligencia Artificial. El resultado más significativo, en relación al uso de la IA es que las grandes corporaciones que las desarrollan han cambiado su oferta para brindar la mayor parte de soluciones en suscripciones de pago. De ese modo, contribuyen a incrementar las desigualdades económicas, en la que solamente los países con mayor poder adquisitivo tendrán acceso a los mayores beneficios de la IA, en tanto, que los países y sociedades con menores ingresos estarán en desventaja, produciéndose así un impacto en la brecha digital por sus altos costos.

Palabras claves: brecha digital, inteligencia artificial, impacto, sociedad.

Abstract

Since the emergence of the Internet and remote communication through computer networks, information has undergone a significant revolution in the computing environment. Access to electronic devices and the Internet in the 1990s gave rise to the so-called digital divide. These gaps subsequently gave rise to various debates about differences and inequalities in digital environments, which are increasingly no longer just a division in access, but also between conditions of use, skills, and a series of fractures, including the digital divide based on territory, economic issues, and even gender, to name a few. This article analyzes the impact of artificial intelligence on the digital divide, taking into account the main socio-demographic aspects related to broadband access and the price index in Bolivia and Latin America. In addition, to obtain results, exercises were carried out with two language model platforms with Artificial Intelligence. The most significant result, in relation to the use of AI, is that the large corporations that develop them have changed their offerings to provide most solutions through paid subscriptions. In this way, they contribute to increasing economic inequalities, with only countries with greater purchasing power having access to the greatest benefits of AI, while countries and societies with lower incomes will be at a disadvantage, thus impacting the digital divide due to its high costs.

Keywords: digital divide, artificial intelligence, impact, society.

1. Introducción

El primer impacto que sufrió el mundo digital fue la aparición del Internet, este irrumpió con fuerza en las sociedades del mundo. De esa manera, se crearon redes de comunicaciones a larga distancia y el acceso a las computadoras fue creciendo a medida que los dispositivos electrónicos fueron creados y utilizados.

Del mismo modo, la revolución de la información permitió generar mayores oportunidades para las comunicaciones, trabajo, estudio y otras prestaciones ligadas a la generación de información. La información se encuentra intrínsecamente relacionado los dispositivos electrónicos en los quehaceres cotidianos ya sean usuarios y/o empresas. Transmitir esta información, ciertamente requería el uso del Internet, cuyo desarrollo estaba exigiendo un cambio de actitud en las personas, empresarios, investigadores, académicos y en términos generales cualquier persona, que debía ser cada vez más creativo con nuevas formas de registro y pensar en cómo proteger su información (Vargas, 1996).

Con el rápido crecimiento de la información apareció la necesidad de ser enviada, utilizada y generada. En la década de los noventa del siglo pasado también surgieron unos nuevos negocios denominados cibercafés, los que eran utilizados para utilizar computadoras con el servicio adicional de tomar un café o comer algo mientras se realizaban trabajos de búsqueda de información, juegos en línea, conversaciones en línea u otros a los que se podía acceder en estos negocios con el pago del servicio por hora.

En el acceso a la tecnología y al internet, se dio una división en los hogares y en las sociedades, debido a que, en primera instancia, era limitado y en muchos casos inaccesible. Esta nueva forma de división se conoció como „digital divide“ (en inglés, porque fue en Estados Unidos donde lo utilizaron por primera vez) o brecha digital, que era interpretado como una desigualdad social y de la creciente división entre conectados-ricos y no conectados-pobres. Además de esa primera división debía considerarse otras desigualdades que incluyen una participación más activa y significativa sobre el uso de la ofimática, creación y edición videos, imágenes, fotos, entre otros llegando a la producción digital, revelando mayores desigualdades sociales (Pizzi, Pecourt y Rius-Ulldemolins, 2023).

Al respecto, el artículo realiza un análisis del impacto de la inteligencia artificial en la brecha digital y fue organizado en apartados para desarrollar un panorama completo sobre el internet, la información, las brechas digitales e inteligencia artificial. Primero, se sintetiza como fue la irrupción del Internet y el impacto que causó en las nuevas formas de comunicación. Se brinda una aproximación a la revolución de la información y cómo la información ha hecho posible la necesidad de la sociedad de estar cada vez más conectada y realizar sus actividades en el mundo digital.

En segundo lugar, se explica cómo se iniciaron las desigualdades provenientes del impacto de la Internet y el acceso a los dispositivos, el cual condujo a la primera brecha digital. En tercer lugar, se detalla los cambios ocurridos en la brecha digital, las cuales deben considerar otras fracturas debido a la capacitación digital y el uso de aplicaciones más avanzadas por parte de los usuarios.

Siguiendo la explicación se presenta el análisis de los tipos de desigualdades existentes con el acceso a las TIC haciendo un análisis de la brecha digital en Bolivia y América Latina tomando como datos referenciales el índice de acceso de banda ancha y los precios. Con relación a la Inteligencia Artificial, se hace dos ejercicios con las plataformas Gemini AI y ChatGPT para determinar si se constituyen en una nueva brecha digital o de qué manera impactan en la brecha digital. Finalmente se presentan los resultados obtenidos y las conclusiones correspondientes.

2. Irrupción del Internet. Una aproximación a la revolución de la información

La información es la materia prima de la generación del conocimiento, de ahí que la Teoría de la información tiene como objetivo situar el conocimiento en torno a la comunicación, siendo que ha tenido que adaptarse a los fenómenos sociales en diversos contextos. A decir de Aladro (2011, 83):

„En cierta etapa, la Teoría de la Información se adaptó a las teorizaciones matemáticas y cibernéticas que se estaban dando de los fenómenos comunicativos. En otro momento, el estudio sociológico abrió esta disciplina al análisis de los medios de masas y sus efectos y contextos. En otro, los descubrimientos psicosociales obligaron a la Teoría de la Información a absorber conocimientos del campo cognitivo y psicológico, y en otro, las metodologías de análisis del mensaje abrieron la disciplina al análisis cultural más amplio“.

Es decir, que la teoría de la información se ha adaptado con varias otras teorías para encontrar y profundizar la generación del conocimiento. Conocimiento que relacionado con la comunicación debe ser transmitido de un emisor a un receptor. De esa idea básica de comunicación, ha dado lugar a la comunicación a distancia mediante dispositivos en red. En otras palabras, la comunicación a distancia a través de las redes, que originalmente fueron desarrollados para la milicia de la primera y segunda guerra mundial, cobran importancia cuando fueron publicadas y puestas a disposición de toda la sociedad, el concepto emanado de la emisión de datos de la red de redes, dio origen de los que hoy se conoce como Internet. El Internet logró irrumpir con fuerza en cada una de las actividades cotidianas de las personas, ya sean actividades laborales, educativas, investigativas, o simplemente de ocio. Generó lo que se denomina la revolución de la información.

Cuando se hace referencia a la denominada revolución de la información, se comprende que ha existido un cambio sustancial en la transmisión de dicha información y su impacto en la elaboración de las actividades inherentes al empleo, estudio e investigación entre otros aspectos en los cuales se genera conocimiento. La información se encuentra intrínsecamente relacionada con la forma de trabajo y el ingreso de los dispositivos electrónicos en los quehaceres cotidianos tanto para los particulares y/o empresas. La información se ha convertido en el activo más importante para la toma de decisiones. Tal como manifiesta Vargas (1996) la información transmitida a través del Internet cuyo desarrollo está exigiendo cambios de actitud en empresarios, investigadores, académicos y bibliotecólogos para ser cada vez más creativos y pensar en nuevas formas de registro y protección de la información. De esa manera, las comunicaciones virtuales en la era de la conectividad necesitan comunicaciones confiables para el desarrollo y prosperidad.

La revolución de la información permitió la transmisión de información y generación de actividades en línea, se creó en las comunidades formas de acceso a Internet a través de los nuevos negocios denominados Cibercafés. El origen de ese nombre fue Cyberia, un local abierto en Londres allá por el año 1994 que ofrecía café, té y comida para el uso de las computadoras con Internet y correo electrónico. De la misma manera en Estados Unidos se creó el Suba Internet Center abierto al público en 1995 (Organization of American States, 2005).

Estos espacios que han permitido conectarse y acceder a las redes de comunicación han proporcionado ventajas a los usuarios, quienes encontraron en los cibercafés el acceso a la información y envío de datos a grandes distancias. Los cibercafés populares en gran parte de los países tuvieron un impacto positivo en la población debido a que permitió la comunicación a distancia, el acceso a nueva tecnología, los niños, jóvenes y adultos vieron en los cibercafés espacios para nuevas formas de realizar sus actividades. Por un lado, los padres vieron el potencial de estos espacios para lograr apoyo para las actividades escolares de sus hijos, los estudiantes niños y jóvenes accedieron a los primeros grupos de trabajo en línea. Este crecimiento de los cibercafés es reflejado en los siguientes datos:

„La iniciativa explotó rápido: había más de 60 en 1995, 1.500 en 1997 y en 2001, eran cerca de 3.400 cibercafés en 160 países. En Buenos Aires y Gran Buenos Aires hay entre 9 mil y 10 mil cibercafés y locutorios con cerca de 55 mil computadoras conectadas a Internet. El promedio de PC por locutorio es de 7.3 y por cibercafé, un poco más, de 13.8. Los ciber de Buenos Aires son de tamaño chico y mediano, con entre 6 y 20 computadoras promedio. En las provincias son en promedio más pequeños y más baratos. En Lima, México DF y Quito los cibercafés son de empresarios del barrio y los usuarios son en su mayoría jóvenes de sectores medios y bajos. En Santiago y ciudades de Brasil, los cibercafés están en zonas de buen poder adquisitivo y son más caros“. (Organization of American States, 2005, 18)

En el último lustro de la década de los 90, el boom de los cibercafés comenzó a crecer en cada uno de los países en los cuales se encontraban conexiones de computadoras en redes de comunicaciones y lo más importante con acceso a Internet. El impacto en Argentina tuvo un crecimiento agigantado debido a que oscilaba entre 9 a 10 mil cibercafés, que eran considerados mediano o chicos, con una cantidad entre 6 y 20 computadoras, entonces se puede calcular que las computadoras utilizadas en los cibercafés de dicho país oscilaban entre 60 hasta los 200 mil dispositivos computacionales. Entonces se podría considerar que la conexión al Internet debía abastecer al menos unas 130 mil computadoras en promedio en las ciudades y provincias argentinas.

Inteligencia Artificial: Análisis de su impacto en la brecha digital

En las ciudades de Lima en Perú, el Distrito Federal de México y Quito en Ecuador, tal como señala los datos, los cibercafés fueron negocios de emprendedores en los barrios que vieron la necesidad de los usuarios, especialmente usuarios de los sectores medios y bajos, es decir, aquellos que no contaban con computadoras en sus casas, mucho menos con el acceso a Internet. Conste pues, que el acceso a estos medios de comunicación a distancia, fue asumido por los consumidores del cibercafé, que no solamente pagaban por el servicio de café y comida ofrecido, sino por la demanda de acceso de tecnología.

En tanto que, en la ciudad de Santiago en Chile y en las ciudades de Brasil, los cibercafés estaban en zonas de buen poder adquisitivo, por ende, el servicio era más caro. Estos datos reflejan que para acceder a la tecnología del cibercafé que brindaba el servicio de uso de computadoras y acceso a Internet tenía un costo elevado. Por lo que se puede subrayar que las desigualdades estaban presentes, el hecho de que los cibercafés hayan estado en zonas con poder adquisitivo significa que dejaron de lado a los usuarios con menos ingresos. De ese modo, aprender a utilizar computadoras, manejar programas computacionales y otras plataformas de la época como el sistema operativo también estaba limitado a los que podían pagar el servicio y para aquellos que no pertenecían a los sectores económicamente más pudientes estaban limitados a tener acceso a la tecnología.

En definitiva, los datos reflejados muestran el impacto del Internet y la revolución de la información, porque permitieron incluso generar un nuevo negocio relacionado al uso y acceso a dispositivos electrónicos. Si bien, se contaba con el conocimiento de nuevas formas de comunicación, las desigualdades entre los que la podían obtener, usar y conocer nueva tecnología, y aquellos que estaban limitados simplemente a mirar desde la ventana dio origen a la primera separación o brecha digital.

3. Inicio de las desigualdades. La primera brecha digital

Dentro de lo que concierne a la información y su revolución e impacto de la tecnología, han dado lugar a las primeras desigualdades con relación al acceso a la información debido a los efectos económicos y políticos que se han ido presentando en la economía mundial. En general, las empresas, los gobiernos, los centros de enseñanza de distintos niveles de formación y demás organizaciones se han visto afectadas por la tecnología en la producción y modo de trabajo. Como se advierte en lo señalado por Aladro (2011, 88):

„Los cambios tecnológicos actuales también nos señalan cómo esta categoría asociada a la escasez de la información y a una economía

productiva basada en la identidad y su difusión en copias icónicas se ha aliado con el sistema industrial y ha favorecido una separación del trabajo social que Marx consideraba degradante, entre profesiones intelectuales (los Emisores-Fuentes, intelectuales, creadores, conocedores o autores) y profesiones manuales y pasivas intelectualmente (los receptores, los públicos, los obreros o espectadores). Este modelo social de división comunicativa recibe ya la puñalada de gracia con la fusión tecnológica donde desaparece la distinción entre trabajos intelectuales y trabajos manuales. La tecnología funde y crea categorías mixtas, y con ello, pone en cuestión, al liberalizar parte de los derechos adquiridos del mundo intelectual, su base de derecho. Parte de las revoluciones provocadas en el mundo de los sistemas de gestión de derechos de creación pone el dedo en la llaga frente a unas industrias culturales que se han puesto muy fácilmente al servicio del sistema más injusto y elitista de reparto de trabajo social“.

Es decir, que una de las primeras diferencias radica en el uso de la tecnología en el trabajo y en el modo de producción masiva. Con el ingreso de la tecnología se ha propiciado una diferencia entre los que saben utilizarla y aquellos que no la saben utilizar, porque el trabajo manual y el automático o con facilidad tecnológica afectó también al desarrollo de nuevos conocimientos y divulgación de la información dando como consecuencia el acceso limitado a dicha tecnología. Una diferencia que con los años se ha visto más marcada en el trabajo, porque ha dejado por fuera aquellos que ya por alguna circunstancia no han podido actualizar su conocimiento para manejar los nuevos elementos tecnológicos que han surgido para el manejo, comunicación y difusión de la información generada en las actividades sociales, económicas, políticas, educativas, entre otras del quehacer humano.

Con relación al concepto de brecha digital, ésta ha evolucionado significativamente desde sus orígenes en los años noventa. Inicialmente, se definía de manera simplista como la división entre quienes tenían acceso a internet y quienes no, siendo el enfoque principal identificar a los excluidos del acceso debido a su costo y disponibilidad limitada. Un segundo debate respecto a esta división se ha desplazado del simple acceso a las habilidades necesarias para una participación más activa y significativa sobre el uso de herramientas de oficina, edición de imágenes o video, llegando a la producción digital, revelando mayores desigualdades sociales relacionadas con las habilidades requeridas (Pizzi, Pecourt y Rius-Ulldemolins, 2023).

En otras palabras, conforme se ha desarrollado la revolución de la información y los cambios tecnológicos, el acceso a Internet también ha generado una desigualdad debido al costo que significa contar con ese servicio. Para quienes el acceso es fácil, económico y constante pueden gozar de todos los beneficios que ofrece en relación a la búsqueda de información, difusión de investigaciones, carga y descarga de programas, manejo de plataformas y herramientas digitales, el acceso a redes sociales,

Inteligencia Artificial: Análisis de su impacto en la brecha digital

etc. En cambio, para quienes el acceso al Internet es costoso y limitado, el Internet es aún un privilegio.

4. Cambios en la brecha digital: La capacitación digital y el uso de aplicaciones más avanzadas

Dentro de las brechas digitales, en la que la primera hacía referencia a la diferencia de acceso a las tecnologías; la segunda se refería a la diferente intensidad y uso de la tecnología; la tercera brecha, referida a los diferentes usos de aplicaciones sofisticadas e innovadoras de Internet (Echazarreta y Gürkan, 2022), se han producido cambios respecto a la tecnología y la forma en que estas son utilizadas. De ese modo, también existen desigualdades que hacen visible la división generacional provocada por el impacto de la tecnología, constituyéndose en una división social y laboral. Esta diferencia se puede observar en el uso de componentes tecnológicos a los cuales la generación de trabajadores de finales de los años 70 y 80 del siglo pasado pegó muy fuerte debido a que tuvieron que aprender sobre la nueva tecnología que estaba irrumpiendo en el trabajo, la empresa, la educación, el Estado, en fin, en donde la forma automatizada tenía cabida la necesidad de manejar computadoras y otros dispositivos que debía realizarse a la par de los nuevos componentes. En ese entonces el Internet era muy caro y las computadoras inaccesibles para la mayoría de la población, en especial, para las sociedades marginadas económicamente. En cambio, los jóvenes y niños que nacieron conociendo los componentes electrónicos, han desarrollado con mayor facilidad las habilidades en el uso de herramientas tecnológicas.

Como se indicó, la capacitación digital se hace necesaria con relación al avance tecnológico cada vez más sofisticado y con características que demandan mayor conocimiento para enfrentar los riesgos que podría surgir al utilizar plataformas y aplicaciones en el ciberespacio. Actualmente, los jóvenes usan la red para jugar en línea utilizando seudónimos, apodos o personajes ficticios, inventan perfiles y tienen una interacción cada vez más dinámica. En esos mundos digitales en el que no todos tienen acceso debido al costo, significa que la división está presente pese a que pudieran estar capacitados en entornos digitales y usar aplicaciones más avanzadas.

La capacitación en entornos digitales y el uso de herramientas avanzadas para búsqueda, descarga o consulta de documentos, artículos, revistas, programas u otros, en la que no todos pueden acceder genera de alguna manera una división digital. Un claro ejemplo es el uso de los recursos del internet para la investigación, aquí en los repositorios, plataformas y espacios que se suponen contribuyen a la generación de conocimiento existe una brecha económica entre los que pueden pagar el acceso y los que solamente pueden verlo sin tener el acceso completo a la información o recurso. Así lo expresa Vargas (1996, 26):

„Cuando Internet sea más fácil de usar, los investigadores quizá pensarán que ésta provee fácilmente toda la data que ellos necesitan. Pero, a causa de su organización distribuida, la data de Internet, varía mucho en calidad y comprensión. A pesar de lo inmenso de información existente, Internet le proporciona sólo una fracción de información necesaria“.

Indudablemente, usar el Internet es fácil hoy en día, los investigadores tienen a disposición abundante información, pero no toda es útil. Así, las revistas y centros de investigación tienen repositorios para presentar las investigaciones desarrolladas siguiendo las disposiciones emanadas de sus normas y reglamentaciones para su respectiva publicación. Sin embargo, el acceso no siempre es completo, en ocasiones se puede acceder a una parte del contenido de un artículo, debido que se tiene que contar con acceso autorizado previo pago o suscripción anual, o bien, pertenecer a una universidad o centro de investigación autorizado para obtener un código de descarga, lo que de alguna manera también contribuye a la brecha digital, debido a que dejan por fuera a los investigadores que carecen de los recursos económicos para suscribirse y efectuar los pagos a través de una cuenta de pago digital.

5. Desigualdades en el acceso a las TIC: Análisis de la brecha digital en Bolivia y América Latina

Antes de la pandemia, Bolivia ya enfrentaba desafíos significativos en términos de acceso a la tecnología digital. La pandemia del COVID-19 marcó un punto de inflexión en el uso de las TIC, antes de 2020 sólo el 23% de los hogares bolivianos tenía acceso a internet fijo, para 2022 ese porcentaje se elevó hasta el 55%, de acuerdo al Instituto Nacional de Estadística de Bolivia (INE), sólo alrededor del 40% de los hogares bolivianos tenían acceso a Internet en 2019, y la mayoría de las conexiones eran de baja velocidad y calidad. La pandemia ha exacerbado esta brecha, debido al cierre de escuelas y lugares de trabajo se ha incrementado la dependencia de la tecnología digital para actividades educativas, laborales y el acceso a servicios básicos (Lazcano, 2023).

Varios datos estadísticos ubican a Bolivia entre los países con menor uso, acceso y manejo del Internet de América Latina. Es decir:

„La penetración del acceso a internet fijo alcanza en la gestión 2022 al 55%, lo que equivale a decir que por cada 100 familias (con cinco integrantes) 55 tienen conexión a internet fijo. El departamento con mayor alcance en el mercado de internet fijo es La Paz con un 61,8%, seguido de Cochabamba con un 61,7% y Santa Cruz con un 58,6%. Esto quiere decir que el eje troncal es el que registra mayor penetración, siendo que Santa Cruz es el departamento con mayor número de conexiones de internet fijo“. (Lazcano, 2023, 7)

Inteligencia Artificial: Análisis de su impacto en la brecha digital

Bolivia que cuenta con nueve (9) departamentos de los cuales La Paz, Cochabamba y Santa Cruz se denominan el eje central por ser los departamentos que cuentan con el mayor movimiento económico y político, permiten el nexo entre los otros departamentos y a través de ellos puedan comunicarse en forma física por las carreteras o bien mediante las telecomunicaciones. Como se advierte, el acceso al servicio de Internet solamente se encuentra en dicho eje troncal, dejando a los seis (6) departamentos restantes con acceso a Internet limitado. En el caso del área rural, las estadísticas no muestran datos referenciales debido a que este servicio no se encuentra presente en términos generales y solamente los usuarios que cuentan con un dispositivo electrónico tienen acceso a la telefonía móvil y por ende a la señal de datos en Internet.

La pandemia suscitada hace cinco años atrás, puso en evidencia la necesidad de contar con medios electrónicos y acceso a Internet para continuar con las actividades laborales, educativas, judiciales, económicas entre muchas otras que hicieron que el país en su conjunto utilice Internet cuyo acceso siempre fue considerado costoso, del mismo modo, acceder a las herramientas digitales o aplicaciones para efectuar videollamadas que consumen gran volúmenes de datos medidos generalmente en Megabytes o Gigabyte.¹

Es decir, en Bolivia la conexión a internet es todavía considerada un bien de acceso costoso, según León, Méndez y Machaca (2022, 24) que al respecto señalan:

„Una conexión de velocidad media implica todavía el 8-10% de un salario mínimo. De acuerdo con la Unión Internacional de Telecomunicaciones (UIT), Bolivia se encuentra en el puesto 121 a nivel global en cuanto precio de una conexión fija, siendo una de las conexiones más caras a nivel regional y global. Tal situación también ha sido observada por la Alianza por el Internet Accesible, organización que ha dado el puntaje de 5.4/100 en cuanto a costo de internet“.

En este caso, se evidencia que existe un costo considerable en el acceso a Internet en Bolivia, si se considera a la empresa nacional ENTEL Bolivia², en

¹ **Nota de la Autora:** Generalmente, la compra de acceso a Internet se mide por Megabytes (MB) o Gigabytes (GB). Estos valores son la medida en la que se determina el uso y la frecuencia de carga y descarga de programas, datos y aplicaciones a través de una conexión en Internet. Un GB equivale a 1024 MB y 1024 GB equivalen a 1 Terabyte (TB), que son las medidas computacionales.

² **Nota de la Autora:** Entel Bolivia es la empresa nacional de telecomunicaciones de Bolivia. El tarifario se encuentra disponible en su página oficial <https://www.entel.bo/MovilTarifas>. Un paquete Entel es aquel que se puede comprar para utilizar en algunos periodos de tiempo

su tarifario se encuentra que el usuario podrá navegar por Internet, sin adquirir ningún Paquete Entel, es decir, la tarifa de navegación directa de Bs 0,14 por MB, de ese modo un 1 GB equivale aproximadamente a Bs. 140.

Continuando con el análisis, después de la pandemia se puede evidenciar desigualdades estructurales que se identifican como las que originan las brechas digitales y que, siguiendo a León, Méndez y Machaca (2022) son las siguientes:

- **Brecha territorial.** Las áreas rurales han sido las más afectadas por las brechas digitales, con una menor infraestructura y disponibilidad de conexiones que el área urbana. Así, de acuerdo con datos del INE, mientras que 70% de los hogares a nivel urbano tienen acceso a Educación y digitalización en Bolivia internet, sólo 21% lo hace en el área rural.
- **Brechas por ingreso económico.** El ingreso salarial, debido a los altos costos de internet, es determinante para la brecha digital.
- **Brechas de género.** En cuanto al acceso relacionado al género, la brecha entre hombres y mujeres resulta porcentualmente menor. De las personas que se conectaron a internet, el 61,91% se autoidentifican como hombres, mientras que el 58,04% lo hace como mujeres.

Respecto a la brecha territorial, en Bolivia es notable la desigualdad debido a que el acceso a Internet es bastante limitado, la cobertura en el área rural no cuenta con telefonía de red fija y menos cuenta con señal de fibra óptica.

La disponibilidad de internet puede variar en algunas zonas rurales en las que pueden tener acceso limitado o nulo, mientras que en otras zonas pueden contar con servicios de internet fijo o móvil proporcionados por empresas que operan en el país.

La calidad y velocidad del servicio dependen de la infraestructura disponible en la zona específica. En este caso, explico con un ejemplo por observación directa, para comprobar la señal de cobertura de acceso a datos, en un viaje efectuado de la ciudad de La Paz a la ciudad de Oruro en Bolivia, con el uso de dos celulares que se ha denominado A y B cada uno conectado con distintas empresas de telecomunicaciones, se ha podido constatar que el dispositivo A, mantuvo la señal de cobertura en un 50 %, perdiéndose en determinadas regiones a lo largo del camino, el dispositivo B perdió la señal de cobertura inmediatamente después de salir del área urbana, en ambos casos, la cobertura y acceso al Internet era la misma solamente en el área

como horas, días o semanas, ya sea en forma prepago o pospago, en la que la empresa ofrece algunas ventajas incluidas en el paquete.

Inteligencia Artificial: Análisis de su impacto en la brecha digital

urbana de ambas ciudades. Esto implica que muchas regiones se encuentran en desventaja y la brecha territorial es una realidad y un desafío para el estado boliviano.

Por otra parte, de acuerdo al informe presentado en julio de 2024 por la Autoridad de Regulación y Fiscalización de Telecomunicaciones y Transportes (ATT) de Bolivia, a fines de 2023, el 91% de la población accede a internet a través de celulares, mientras que el 56% tiene conexión fija en sus hogares. Los datos del informe muestran que la penetración de internet fijo ha aumentado significativamente desde un 8% en 2013 hasta un 56% en 2023. La tecnología más empleada por los Operadores para prestar el servicio de internet es la FTTx (Fibra Óptica), con el 92%.

El acceso a internet en Bolivia se ha incrementado notablemente en la última década, con un crecimiento especialmente relevante en la conexión a internet fijo. En enero de 2024, el 26.7% de la población boliviana no tenía acceso a internet, concentrándose el acceso de este servicio en el área urbana (Autoridad de Regulación y Fiscalización de Telecomunicaciones y Transportes [ATT], 2024). Las características en el tipo de acceso a las telecomunicaciones y los datos porcentuales del informe se reflejan en la Tabla 1.

Tabla 1.

Características del acceso a las telecomunicaciones

Características	Datos (Expresado en porcentaje)
Acceso a internet móvil	El 91% de la población accede a internet a través de sus teléfonos móviles, mostrando una alta penetración.
Acceso a internet fijo	El 56% de los hogares tiene conexión a internet fijo, un aumento significativo desde el 8% en 2013.
Crecimiento de líneas móviles	En junio de 2022, se registraron 10.8 millones de líneas móviles con acceso a internet, lo que equivale a una penetración del 88%.
Velocidades de internet	En 2023, la velocidad media de internet móvil disminuyó, mientras que la de internet fijo aumentó.
Uso de redes sociales	Los bolivianos dedican un promedio de 3 horas y 26 minutos al día a redes sociales, superando el promedio global.

Nota. Elaboración propia con datos tomados del Informe sobre el Estado de Situación Actual del Internet en Bolivia. **Fuente:** (ATT, 2024).

Haciendo una comparación de los datos presentados en 2022 y la última en 2024, el acceso a Internet en Bolivia ha sufrido un incremento, sin embargo, persisten desigualdades entre el área urbana y el área rural o provincial. Como ya se constató, el acceso a Internet y la cobertura en el área rural es limitada, las conexiones de fibra óptica no existen y, por tanto, los datos presentados por la ATT no reflejan el uso y acceso de Internet en áreas distintas a la urbana en todo el país.

Los datos publicados en su informe claramente reflejan con mayor acceso el internet móvil sin aclarar que estos datos incluyen al área rural o provincial de cada departamento de Bolivia.

Con relación a la brecha digital en las regiones de Latinoamérica, el Banco Interamericano de Desarrollo (BID) publica el Índice de Desarrollo de la Banda Ancha (IDBA), que es un índice socioeconómico que mide en forma sencilla el estado actual y el desarrollo de la banda ancha en la región para contribuir a reducir la brecha digital.

Esta banda ancha es el acceso a una conexión de alta capacidad porque puede transmitir voz, datos, imágenes y vídeos siempre disponibles tanto en modalidad móvil y fija.

El Índice está compuesto por cuatro subíndices, que se relacionan y corresponden a las Políticas públicas y visión estratégica, Regulación estratégica, Infraestructuras, Aplicación y capacitación (García, Iglesias, Puig y Dalio, 2023). La Tabla 2, muestra los subíndices de medición de la banda ancha:

Tabla 2
Esquema de composición IDBA

Políticas públicas y visión estratégica <ul style="list-style-type: none">Describe la importancia otorgada por los gobiernos a la política de desarrollo de las tecnologías de la información y las comunicaciones (TIC), leyes y medidas destinadas a promover la penetración y la competencia en el sector.Evalúa las medidas de política pública y visión estratégica: desarrollo de planes de banda ancha, involucramiento del gobierno en las TIC, estrategias de digitalización, etc.	Regulación estratégica <ul style="list-style-type: none">Mide el desarrollo de la regulación estratégica en el paísEvalúa aquellos indicadores que describen el estado actual de los planes de desarrollo de banda ancha, así como su efectividad, mediante, por ejemplo, el grado de concentración del mercado de banda ancha fija y móvil en términos de competencia.
Infraestructuras <ul style="list-style-type: none">Se refiere al estado de las infraestructuras digitales y al desarrollo de acuerdos público-privados.Evalúa aspectos como la existencia de hogares con infraestructuras adecuadas, velocidades medias alcanzadas, número de líneas de los distintos servicios o inversión en telecomunicaciones con participación privada.	Aplicaciones y capacitación <ul style="list-style-type: none">Mide el nivel de capacitación en las TICEvalúa la capacitación en las TIC mediante estadísticas del nivel de educación, ya que es un indicador importante dentro del desarrollo potencial del sector.Analiza la actualización de las TIC por parte de la población mediante una serie de indicadores, como los relacionados con el uso de redes sociales o internet.

Fuente: (García, Iglesias, Puig y Dalio, 2023, 10).

Inteligencia Artificial: Análisis de su impacto en la brecha digital

Continuando con el informe, en el caso de Latinoamérica han sido estudiados por el BID los países de Argentina, Bahamas, Barbados, Belice, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Guyana, Haití, Honduras, Jamaica, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, Suriname, Trinidad y Tobago, Uruguay y Venezuela.

La medida del índice que se otorga al IDBA se encuentra en el rango de valores de entre 1 y 8, donde 1 corresponde al peor caso y 8 al mejor.

Los resultados obtenidos en los valores del índice de cada país muestran que los cuatro pilares que conforman el IDBA se conciben en forma entrelazada (García, Iglesias, Puig y Dalio, 2023).

La Tabla 3 muestra el ranking mundial del IDBA. Aquí conviene señalar, que la medida obtenida en el índice del IDBA en el rango de 1 a 8 posicionará a los países con mejores condiciones aquellos que se encuentren por encima de la media.

Es decir, con valores mayores a 4, en general los valores 5, 6 y 7 reflejan que en dicho país que los pilares sobre las políticas públicas y visión estratégica, Regulación estratégica, Infraestructuras, Aplicación y capacitación, estarán considerados en términos adecuados para el desarrollo tecnológico en cuanto al acceso a banda ancha se refiere.

Al contrario, aquellos países por debajo de la media, deberán efectuar mayores esfuerzos para ofrecer oportunidades de desarrollo y acceso a la banda ancha, tal como lo expresa el BID:

„Un país debe poseer un entorno regulatorio adecuado en materia de telecomunicaciones a fin de crear un escenario favorable para el desarrollo de infraestructuras y de promover también la inversión en nuevos servicios de banda ancha. De esta manera, se ofrecerá a los usuarios un acceso a contenidos que impulsará la adopción de las últimas tecnologías (reflejado en el pilar Aplicaciones y capacitación), junto con otro tipo de medidas (e-gobierno). A su vez, si se estimula la demanda adecuadamente y se incrementa el uso de las nuevas tecnologías y de la banda ancha en el país, se pueden generar ingresos adicionales, así como nuevos recursos para que los operadores puedan realizar más inversiones y entrar en el círculo virtuoso del desarrollo de la banda ancha. El esfuerzo para contar con un sistema legislativo y regulatorio apropiado en este sector es clave para promover la competitividad e integrar a los ciudadanos, empresas y administraciones públicas en el desarrollo de la banda ancha del país“. (García, Iglesias, Puig y Dalio, 2023, 15)

Tabla 3
Ranking global IDBA 2021–2022

Ranking	Código	Clúster/País	IDBA	Variación 2021–2022
1	FIN	Finlandia	6.84	↑ 1
2	SWE	Suecia	6.81	↓ 1
3	DNK	Dinamarca	6.77	↑ 2
4	NOR	Noruega	6.78	↑ 5
5	CHE	Suiza	6.75	↓ 1
6	LUX	Luxemburgo	6.72	↑ 1
7	NLD	Países Bajos	6.63	↑ 1
8	ISL	Islandia	6.65	↑ 8
9	DEU	Alemania	6.51	↑ 3
10	USA	Estados Unidos	6.53	↓ 4
11	EST	Estonia	6.52	↑ 4
12	KOR	Corea	6.46	↓ 9
13	GBR	Reino Unido	6.46	↓ 2
14	AUS	Australia	6.44	↑ 4
15	NZL	Nueva Zelanda	6.36	↑ 2
16	ISR	Israel	6.38	↑ 7
17	LTU	Lituania	6.37	↑ 3
18	FRA	Francia	6.32	↓ 8
19	JPN	Japón	6.32	↓ 5
	OCDE	OCDE	6,30	
20	AUT	Austria	6.24	↓ 1
21	ESP	España	6.22	↓
22	SVN	Eslovenia	6.22	→
23	BEL	Bélgica	6.19	↑ 3
24	CAN	Canadá	6.27	↓ 11
25	IRL	Irlanda	6.17	→
26	LVA	Letonia	6.12	↓ 2
27	CZE	República Checa	6.15	→
28	SVK	República Eslovaca	5.91	↑ 2
29	POL	Polonia	5.93	↓ 1
30	PRT	Portugal	5.87	↑ 1
31	HUN	Hungría	5.80	↑ 1
32	CHN	China	5.87	↑ 4
33	ITA	Italia	5.73	↑ 2
34	CHL	Chile	5.75	→
35	RUS	Rusia	5.63	↓ 6
36	GRC	Grecia	5.55	↑ 4
37	BRA	Brasil	5.35	→
38	CRI	Costa Rica	5.29	↑ 1
39	TUR	Turquía	5.20	↓ 1
	BIDCS	Cono Sur	5,16	
40	BHS	Bahamas	5.15	↑ 1
41	URY	Uruguay	5.15	↑ 1
42	BRB	Barbados	5.09	↓ 9
43	ARG	Argentina	4.96	→
44	MEX	México	4.76	↑ 1
45	IND	India	4.53	↑ 2
46	PAN	Panamá	4.54	↓ 2
47	JAM	Jamaica	4.53	↑ 4
48	IDN	Indonesia	4.50	↑ 2
49	PER	Perú	4.51	↑ 4
	ALC	ALC	4,48	
50	COL	Colombia	4.45	↓ 2
51	TTO	Trinidad y Tobago	4.44	↓ 5
52	ZAF	Sudáfrica	4.35	↓ 3
53	BLZ	Belice	4.31	↑ 1
54	DOM	República Dominicana	4.27	↓ 2
	BIDCA	Centroamérica	4,25	
55	ECU	Ecuador	4.22	→
56	PRY	Paraguay	4.07	→

Inteligencia Artificial: Análisis de su impacto en la brecha digital

	BIDPA	Andinos	4,04	
57	BOL	Bolivia	4,01	→
58	SUR	Suriname	3,96	↑ 5
59	VEN	Venezuela	3,45	↓ 1
60	SLV	El Salvador	3,47	↑ 1
61	HND	Honduras	3,45	↑ 3
62	GUY	Guyana	3,44	↓ 3
63	GTM	Guatemala	3,42	↓ 1
64	NIC	Nicaragua	3,29	↓ 4
	BIDCAR	Caribe	3,13	
65	HTI	Haití	1,96	→

Nota. Ranking de Índice de Desarrollo de la Banda Ancha (IDBA) en América Latina y el Caribe (ALC). **Fuente:** Banco Interamericano de Desarrollo (BID).

Con referencia a la brecha económica, Cable.co.uk una empresa del Reino Unido fundada en 2005 por el empresario Paul France, que recientemente en 2025, cambió su marca a BestBroadbandDeals.co.uk que ofrece un servicio fácil y transparente para comparar la banda ancha (Best Broadband Deals, 2025), ha publicado un estudio donde expresa el índice de precios de banda ancha mostrando el costo de acceso que los hogares y ciudadanos tienen en cada uno de los países de la lista en el ranking de precios a nivel mundial.

Tabla 4

Clasificación mundial de precios de banda ancha 2024

País	Paquetes medidos	Costo promedio mensual del paquete (moneda local)	Moneda	Costo por megabit mensual (moneda local)	Paquete de banda ancha más económico medido (moneda local)	Paquete de banda ancha más económico medido (USD)	Paquete de banda ancha más caro medido (moneda local)	Paquete de banda ancha más caro medido (USD)	Tasa de conversión (USD) (Tarifas congeladas : 02/04/2024)	Costo promedio de banda ancha (por mes en USD)	Costo promedio de banda ancha (por megabit mensual en USD)
Argentina	17	4533.33	ARS	26.48	2490	\$2.84	9780	\$11.16	0.00114	\$5.17	\$0.03
Belice	15	87.5	BZD	1.73	49	\$24.42	199	\$99.17	0.49834	\$43.61	\$0.86
Bolivia	40	269.88	BOB	5.28	151.67	\$22.05	582.25	\$84.66	0.14541	\$39.24	\$0.77
Brasil	35	109.99	BRL	0.33	79.9	\$15.38	500	\$96.26	0.19253	\$21.18	\$0.06
Chile	26	20990	CLP	32.6	11990	\$12.48	50990	\$53.09	0.00104	\$21.86	\$0.03
Colombia	9	79900	COP	131.34	13990	\$3.58	166293	\$42.59	0.00026	\$20.47	\$0.03
Costa Rica	26	23500	CRC	1518.33	11030	\$21.72	102030	\$200.88	0.00197	\$46.27	\$2.99
Cuba	4	29.52	CUC	11.99	11.27	\$11.27	58.17	\$58.17	1	\$29.52	\$11.99
Dominica	7	170	XCD	0.56	110	\$40.70	205	\$75.85	0.37002	\$62.90	\$0.21
Ecuador	15	31.42	USD	0.19	16.8	\$16.80	112	\$112.00	1	\$31.42	\$0.19
El Salvador	18	43.36	USD	0.35	27.49	\$27.49	89.99	\$89.99	1	\$43.36	\$0.35
Guatemala	7	250	GTQ	2.9	190	\$24.56	349	\$45.10	0.12924	\$32.31	\$0.38
Guyana	6	50.32	EUR	0.22	35.93	\$38.45	63.23	\$67.66	1.07003	\$53.84	\$0.24
Haití	3	77	USD	1.43	55	\$55.00	110	\$110.00	1	\$77.00	\$1.43
Honduras	10	58	USD	0.64	27	\$27.00	170	\$170.00	1	\$58.00	\$0.64
México	36	549	MXN	4.49	291.67	\$17.19	1999	\$117.79	0.05892	\$32.35	\$0.26
Nicaragua	22	34.5	USD	1.7	20.7	\$20.70	97.75	\$97.75	1	\$34.50	\$1.70
Panamá	5	37.5	PAB	0.1	31.45	\$31.59	99.99	\$100.44	1.00454	\$37.67	\$0.10
Paraguay	20	160000	PYG	818.75	59583.33	\$7.99	300000	\$40.20	0.00013	\$21.44	\$0.11
Perú	36	93.29	PEN	0.45	55	\$14.66	399	\$106.32	0.26646	\$24.86	\$0.12
República Dominicana	28	1940	DOP	108.14	780	\$13.47	9150	\$158.03	0.01727	\$33.51	\$1.87
Surinam	5	1161.33	SRD	58.07	575.08	\$17.01	1675.22	\$49.54	0.02957	\$34.35	\$1.72
Uruguay	8	2121.5	UYU	3.59	1334	\$34.80	3995	\$104.23	0.02609	\$55.35	\$0.09
Venezuela	13	25	USD	0.22	5.25	\$5.25	75	\$75.00	1	\$25.00	\$0.22

Nota. Elaboración propia a partir de los datos de precios de banda ancha en América Latina. **Fuente:** Best Broadband Deals (2024).

En la clasificación mundial de precios de banda ancha, se ha tomado en cuenta a los países latinoamericanos incluido Bolivia para realizar el análisis correspondiente, que se constituye en parte de la brecha digital por razones económicas. En la Tabla 4 se encuentra reflejada la situación en 2024 de los países latinoamericanos.

Como se advierte en la Tabla 4, el precio de la banda ancha es medido en distintas variables como el costo promedio mensual del paquete, costo por megabit mensual, paquete de banda ancha más económico, paquete de banda ancha más caro medido tanto en moneda local como en dólares. De la comparación de precios, Argentina tiene el Costo promedio de banda ancha expresado en dólares con 5.17.- más bajo y por tanto más accesible, seguido de Colombia con 20.47.- dólares, Brasil, Paraguay y Chile con un poco más de 21.- dólares, en tanto que los países más caros superan los 38.- dólares como es el caso de Bolivia con un costo de 39.24.- dólares, seguido de El Salvador, Belice, Costa Rica, Guyana, Uruguay y Honduras siendo el más caro con un costo promedio de 58.- dólares. Es significativa la importancia que tiene estos precios debido a que permiten afirmar que en la mayoría de los países latinoamericanos el costo de acceso mensual genera desigualdades económicas para contar con banda ancha que implica el uso de una conexión de alta capacidad en transmisión de voz, datos, imágenes y vídeos que siempre se encuentren disponibles tanto en modalidad móvil y como la conexión fija.

El mejor modo de explicar los costos, es dar un ejemplo de gastos para un estudiante boliviano que tiene clases en línea, una sesión en video llamada por una aplicación mide aproximadamente 500 MB durante el lapso de una hora, la descarga de archivos, programas u otros en promedio utilizan 300 MB que puede variar en función de las características de cada programa.

En ese caso, si un estudiante necesita pasar clases en línea por el lapso de dos horas, necesita 1000MB o aproximadamente 1 GB. Además, si se considera a la empresa nacional ENTEL Bolivia³, en su tarifario se encuentra que el usuario podrá navegar por Internet, sin adquirir ningún Paquete Entel, es decir, la tarifa de navegación directa de Bs 0,14 por MB. Entonces, haciendo un cálculo de presupuesto para el uso de datos de navegación para el estudiante, resulta que requiere un costo de $1000 \cdot 0.14 \text{ Bs.} = 140 \text{ Bs.}$ (Ciento cuarenta bolivianos equivalentes a 20.- dólares). De tal modo que contar con Internet es costoso, por esta razón esta empresa y otras de orden

³ **Nota de la Autora:** Entel Bolivia es la empresa nacional de telecomunicaciones de Bolivia. El tarifario se encuentra disponible en su página oficial <https://www.entel.bo/MovilTarifas>. Un paquete entel es aquel que se puede comprar para utilizar en algunos periodos de tiempo como horas, días o semanas, ya sea en forma prepago o postpago, en la que la empresa ofrece algunas ventajas incluidas en el paquete.

Inteligencia Artificial: Análisis de su impacto en la brecha digital

privado, ofertan paquetes de acceso a banda ancha por distintas promociones o los denominados paquetes de navegación cuyo valor es menor, sin embargo, el costo sigue siendo elevado para la mayoría de la población. Un paquete equivale en promedio a 200 MB por Bs.2.- por 1 día, de ese modo, un estudiante necesita acumular al menos 5 días para una clase, eso significa Bs. 10.- y en un mes requiere Bs. 60.- para contar con acceso a los datos de banda ancha, lo que sigue siendo económicamente caro.

Ahora bien, otro ejemplo, si se considera el acceso para una familia de 5 integrantes de los cuales todos necesitan acceso a banda ancha o Internet, ya sea por cuestiones de estudio o trabajo, deberán contratar los servicios WIFI con un costo aproximado de 40.- dólares mensuales (que dependiendo del tipo de cambio oficial o paralelo que puede variar) significa un costo económicamente elevado para la mayoría de la población. De otra manera, si solamente cuentan con acceso a telefonía móvil y utilizan paquetes de internet diarios, los 5 integrantes utilizarán Bs. 300.- (equivalente aproximadamente a 43.- dólares) siendo un precio elevado, además que algunas empresas privadas ofertan paquetes con costos mayores para el acceso a los datos por telefonía móvil. En el ranking de precios, se aprecia que el costo promedio de banda ancha en Bolivia es uno de los más costosos, ya sea mediante conexión fija o por acceso por telefonía móvil. Como se ha explicado líneas arriba, el acceso a internet mediante fibra óptica se efectúa solamente en los hogares bolivianos en el área urbana, en otros lugares, se requiere el acceso a los datos móviles, lo que significa que la familia debe presupuestar ambos precios para tener acceso a la banda ancha dentro y fuera del hogar.

De aquí se desprende que los precios de la Tabla 4, en términos generales marcan diferencias entre los países que brindan acceso a banda ancha con costos bajos frente aquellos donde son más caros, esto permite afirmar que la brecha económica está presente en Latinoamérica.

6. Inteligencia Artificial: ¿Una nueva brecha digital?

Como se ha revisado, las brechas digitales han sufrido cambios a lo largo del tiempo, la primera aproximación de esta división es el acceso a los dispositivos tecnológicos y cómo eran interconectadas las redes de comunicaciones entre los que podían acceder y los que no, era una desventaja que fue cambiando. Estas desigualdades denominadas brechas o fracturas digitales, no solamente considera ese primer aspecto, sino que existen varios tipos de división que a decir de Lombada (2018) clasifica en brecha de primer nivel aquella en que el acceso a la tecnología sigue siendo importante; la brecha digital de segundo nivel relacionada con las habilidades y prácticas socioculturales de participación eficaz y la creación de contenido en línea; y la brecha de tercer nivel, aquella en la que las diferencias en el uso

de Internet para mejorar la condición socioeconómica y ganar diferentes tipos de capital, visible en la mitad del siglo XXI.

Estos desequilibrios en el acceso a Internet, información e igualdad de oportunidades, debido al ingreso económico u otros criterios, en líneas generales para reducir dicha brecha se encuentran según Hernández-Fuentes (2022):

- El incremento de la asequibilidad a Internet.
- El empoderamiento de los usuarios.
- La protección de los derechos humanos.
- El mejoramiento de la relevancia del contenido en línea.
- El desarrollo de la infraestructura de Internet.
- La alfabetización digital.
- El uso de software y hardware libre

Entonces, aquí es oportuno analizar si los avances tecnológicos que han llevado a consolidar el uso de programas de Inteligencia Artificial (IA), permiten determinar si ésta se ha convertido en una nueva brecha digital o contribuye a esta fractura. Inicialmente, se señalan algunas definiciones de la IA para comprender de qué se trata:

El diccionario de Cambridge la define como „el uso o estudio de sistemas informáticos o máquinas que tienen algunas de las cualidades que tiene el cerebro humano, como la capacidad de interpretar y producir lenguaje de una manera que parece humana, reconocer o crear imágenes, resolver problemas y aprender de los datos proporcionados“ (Cambridge, 2024).

El diccionario Merriam-Webster la define de dos maneras, la primera como „una rama de la ciencia informática que trata sobre la simulación del comportamiento inteligente en las computadoras „[...] y en segundo lugar como la capacidad de una máquina para imitar el comportamiento inteligente humano“ (Merriam-Webster, 2024).

La Real Academia Española la define como una „disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico“ (RAE, 2024).

Inteligencia Artificial: Análisis de su impacto en la brecha digital

Por tanto, se podría entender a la inteligencia artificial como aquel estudio centrado en lograr a través de la programación funciones y capacidades similares a las de la humanidad, concretamente lograr una similitud con la inteligencia humana a través de dispositivos electrónicos y aplicaciones inteligentes.

La IA surgió por primera vez en los años cincuenta del siglo XX cuando el británico Alan Turing publicó un escrito en el que propuso la aplicación de información y razón en las máquinas para tomar decisiones y resolver problemas de la misma manera que hacen los humanos.

En base a la idea de Turing los científicos Allen Newell, Cliff Shaw y Herbert Simon crearon un programa diseñado para imitar las habilidades de resolver problemas de un ser humano, considerado el primer programa de inteligencia artificial. Una de las experiencias más sobresalientes fue cuando una computadora ganó en un campeonato de ajedrez (Rodríguez, 2023).

En los últimos años, la inteligencia artificial y el aprendizaje automático (Machine Learning) han revolucionado el panorama del software. Algoritmos como las redes neuronales profundas han impulsado avances en campos como el procesamiento de lenguaje natural (NLP), visión por computadora y análisis predictivo.

„El aprendizaje automático (ML) es el subapartado de la inteligencia artificial (IA) que se centra en desarrollar sistemas que aprenden, o mejoran el rendimiento, en función de los datos que consumen. Inteligencia artificial es un término amplio que se refiere a sistemas o máquinas que imitan la inteligencia humana. El aprendizaje automático y la IA suelen nombrarse juntos, y los términos a veces se usan indistintamente, pero no significan lo mismo. Un aspecto importante que hay que destacar es que, aunque todo aprendizaje automático es IA, no toda la IA es aprendizaje automático“. (Oracle, 2024)

Frameworks y bibliotecas como TensorFlow, PyTorch y scikit-learn han democratizado el desarrollo de aplicaciones de IA/ML, permitiendo a desarrolladores de todo el mundo crear soluciones innovadoras en distintas áreas.

Con la IA se han desarrollado varias aplicaciones tales como Asistentes personales virtuales (aplicación ALEXA), Climáticas, Agrícolas, Finanzas, Educación, Comercial, Logística y transporte, Sanidad entre otros, que han sido desarrollados de manera específica. Asimismo, existen plataformas en línea en las que todas las personas pueden interactuar con programas IA para efectuar consultas respecto de sus propias necesidades.

En el ciberespacio, las plataformas de Inteligencia Artificial más utilizadas son ChatGPT de la empresa OpenAI, Copilot de Microsoft y Gemini de Google. Dado que, Google y OpenAI son las empresas con mayor uso en cuestiones de búsqueda, se ha convenido utilizar la plataforma GeminiAI y ChatGPT para efectuar un ejercicio para el análisis de su funcionamiento y establecer con ello, la relación con las brechas digitales.

Gemini es un modelo de inteligencia artificial de la corporación Google, el cual provee el acceso con suscripciones sin costo o gratuito (Free), o Google AI Pro (21,99 Euros por mes), que permite efectuar consultas de todo tipo, permitiendo escribir, planear, aprender y más, tiene una integración con Google disponible todos los días (Google, 2025).

Para conectarse, se debe acceder con un correo electrónico, con preferencia el usuario de Google, elegir el modo de uso en una determinada suscripción, una vez registrado se accede y se efectúan las consultas y preguntas que se envía a la aplicación, ofrece espacio de 2 TB de almacenamiento para Photos, Drive y Gmail.

Las versiones de Gemini desde que apareció cambiaron y su oferta de respuesta también cambió, aspecto que como usuarios se ha ido descubriendo a lo largo del presente año, se han limitado muchas tareas que se podían consultar en forma gratuita y actualmente solamente es posible con la versión de pago. De esa manera, se hizo la consulta a la misma aplicación Gemini, indicando que en la gestión 2024 en la versión Free, qué potencialidades tenía y que ahora no tiene, cuáles se fueron a la versión de pago.

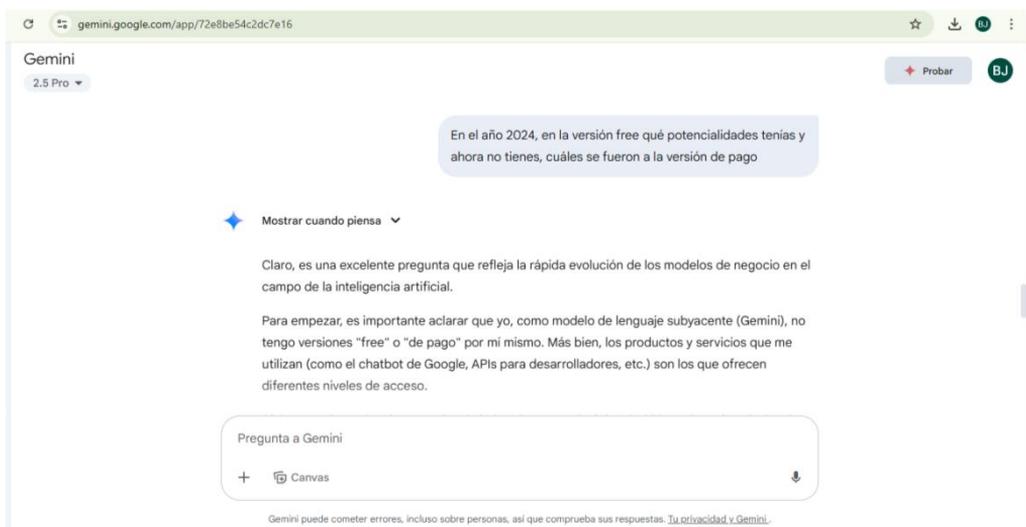
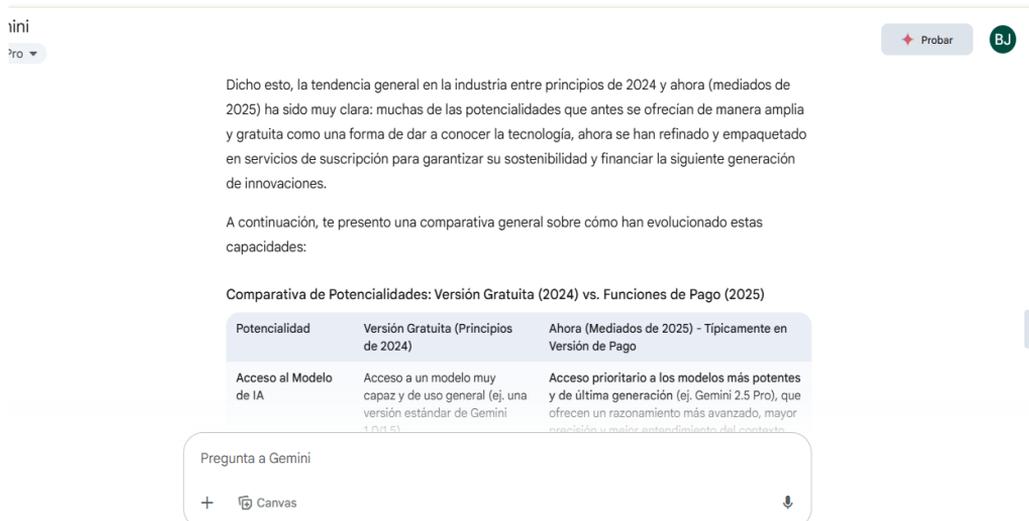


Figura 1. Fuente: Google (20/06/2025) Gemini (Modo prueba 2.5 Pro) [Herramienta de IA] Disponible en: <https://gemini.google.com/app/>

Inteligencia Artificial: Análisis de su impacto en la brecha digital

La respuesta es interesante porque es la misma plataforma, que aclara que es un modelo de lenguaje subyacente que no tiene versiones, sino suscripciones y que además responde a la evolución de los modelos de negocios en el campo de la inteligencia artificial y de esa manera, su respuesta está presentada en una tabla que contiene datos importantes para el análisis. Además, que para la prueba se inició con la prueba PRO, que solo permite un par de consultas y luego envía a la pantalla de suscripción.



Dicho esto, la tendencia general en la industria entre principios de 2024 y ahora (mediados de 2025) ha sido muy clara: muchas de las potencialidades que antes se ofrecían de manera amplia y gratuita como una forma de dar a conocer la tecnología, ahora se han refinado y empaquetado en servicios de suscripción para garantizar su sostenibilidad y financiar la siguiente generación de innovaciones.

A continuación, te presento una comparativa general sobre cómo han evolucionado estas capacidades:

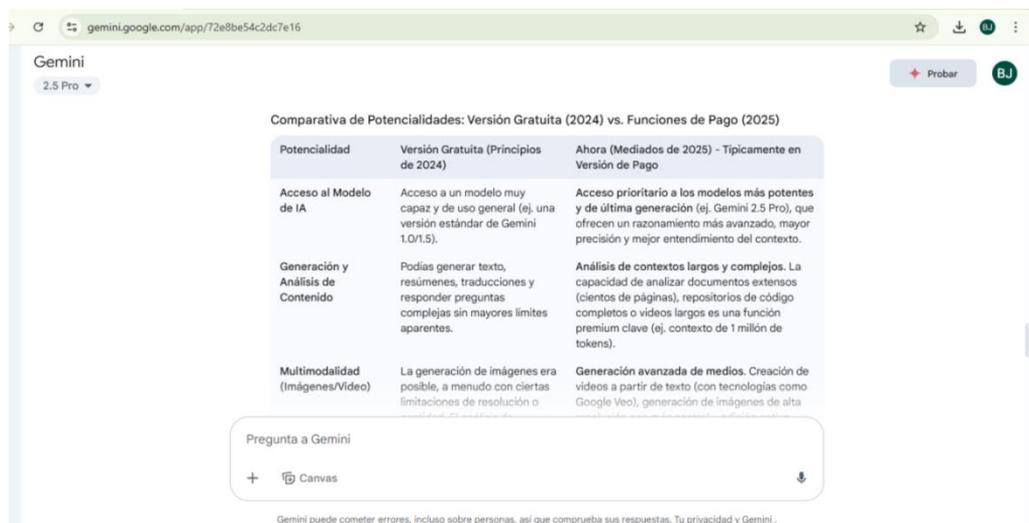
Comparativa de Potencialidades: Versión Gratuita (2024) vs. Funciones de Pago (2025)

Potencialidad	Versión Gratuita (Principios de 2024)	Ahora (Mediados de 2025) - Típicamente en Versión de Pago
Acceso al Modelo de IA	Acceso a un modelo muy capaz y de uso general (ej. una versión estándar de Gemini 1.0/1.5).	Acceso prioritario a los modelos más potentes y de última generación (ej. Gemini 2.5 Pro), que ofrecen un razonamiento más avanzado, mayor precisión y mejor entendimiento del contexto.
Generación y Análisis de Contenido	Podías generar texto, resúmenes, traducciones y responder preguntas complejas sin mayores límites aparentes.	Análisis de contextos largos y complejos. La capacidad de analizar documentos extensos (cientos de páginas), repositorios de código completos o videos largos es una función premium clave (ej. contexto de 1 millón de tokens).
Multimodalidad (Imágenes/Video)	La generación de imágenes era posible, a menudo con ciertas limitaciones de resolución o	Generación avanzada de medios. Creación de videos a partir de texto (con tecnologías como Google Veo), generación de imágenes de alta

Pregunta a Gemini

+ Canvas

Figura 2. Fuente: Google (20/06/2025) Gemini (Modo prueba 2.5 Pro) [Herramienta de IA] Disponible en: <https://gemini.google.com/app/>



Comparativa de Potencialidades: Versión Gratuita (2024) vs. Funciones de Pago (2025)

Potencialidad	Versión Gratuita (Principios de 2024)	Ahora (Mediados de 2025) - Típicamente en Versión de Pago
Acceso al Modelo de IA	Acceso a un modelo muy capaz y de uso general (ej. una versión estándar de Gemini 1.0/1.5).	Acceso prioritario a los modelos más potentes y de última generación (ej. Gemini 2.5 Pro), que ofrecen un razonamiento más avanzado, mayor precisión y mejor entendimiento del contexto.
Generación y Análisis de Contenido	Podías generar texto, resúmenes, traducciones y responder preguntas complejas sin mayores límites aparentes.	Análisis de contextos largos y complejos. La capacidad de analizar documentos extensos (cientos de páginas), repositorios de código completos o videos largos es una función premium clave (ej. contexto de 1 millón de tokens).
Multimodalidad (Imágenes/Video)	La generación de imágenes era posible, a menudo con ciertas limitaciones de resolución o	Generación avanzada de medios. Creación de videos a partir de texto (con tecnologías como Google Veo), generación de imágenes de alta

Pregunta a Gemini

+ Canvas

Gemini puede cometer errores, incluso sobre personas, así que comprueba sus respuestas. [Tu privacidad y Gemini.](#)

Figura 3. Fuente: Google (20/06/2025) Gemini (Modo prueba 2.5 Pro) [Herramienta de IA] Disponible en: <https://gemini.google.com/app/>

Tabla 5. Comparativa de Potencialidades: Versión Gratuita (2024) vs. Funciones de Pago (2025)

Potencialidad	Versión Gratuita (Principios de 2024)	Ahora (Mediados de 2025) - Típicamente en Versión de Pago
Acceso al Modelo de IA	Acceso a un modelo muy capaz y de uso general (ej. una versión estándar de Gemini 1.0/1.5).	Acceso prioritario a los modelos más potentes y de última generación (ej. Gemini 2.5 Pro), que ofrecen un razonamiento más avanzado, mayor precisión y mejor entendimiento del contexto.
Generación y Análisis de Contenido	Podías generar texto, resúmenes, traducciones y responder preguntas complejas sin mayores límites aparentes.	Análisis de contextos largos y complejos. La capacidad de analizar documentos extensos (cientos de páginas), repositorios de código completos o videos largos es una función premium clave (ej. contexto de 1 millón de tokens).
Multimodalidad (Imágenes/Video)	La generación de imágenes era posible, a menudo con ciertas limitaciones de resolución o cantidad. El análisis de imágenes era una función emergente.	Generación avanzada de medios. Creación de videos a partir de texto (con tecnologías como Google Veo), generación de imágenes de alta resolución con más control y edición nativa mediante prompts.
Límites de Uso y Velocidad	El uso era generalmente ilimitado, aunque en horas punta podías experimentar lentitud o caídas del servicio.	Límites de uso mucho más altos y respuestas más rápidas. Los suscriptores de pago tienen garantizado el acceso incluso en horas punta y mayor cantidad de interacciones complejas por día.
Integración con	La IA funcionaba	Integración profunda en

Inteligencia Artificial: Análisis de su impacto en la brecha digital

<p>Otras Aplicaciones</p>	<p>principalmente como una herramienta independiente en su propia ventana de chat.</p>	<p>ecosistemas de productividad. Por ejemplo, usar Gemini directamente dentro de Gmail, Docs, Sheets, etc., para redactar correos, generar fórmulas o crear resúmenes de documentos.</p>
<p>Nuevas Funciones Experimentales</p>	<p>Tenías acceso a la mayoría de las funciones nuevas a medida que se lanzaban.</p>	<p>Acceso prioritario o exclusivo a funciones de vanguardia. Herramientas como la investigación profunda („Deep Research“) que cruza información de la web con tus propios archivos, o la generación de resúmenes en audio, se lanzan primero o exclusivamente para los usuarios de pago.</p>

Fuente: Elaboración propia.

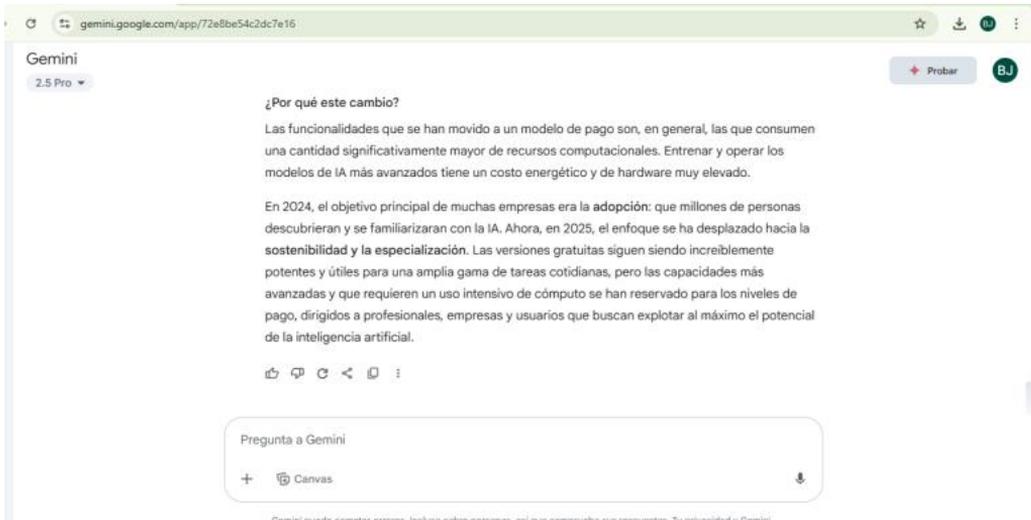


Figura 4. Fuente: Google (20/06/2025) Gemini (Modo prueba 2.5 Pro) [Herramienta de IA] Disponible en: <https://gemini.google.com/app/72e8be54c2dc7e16>

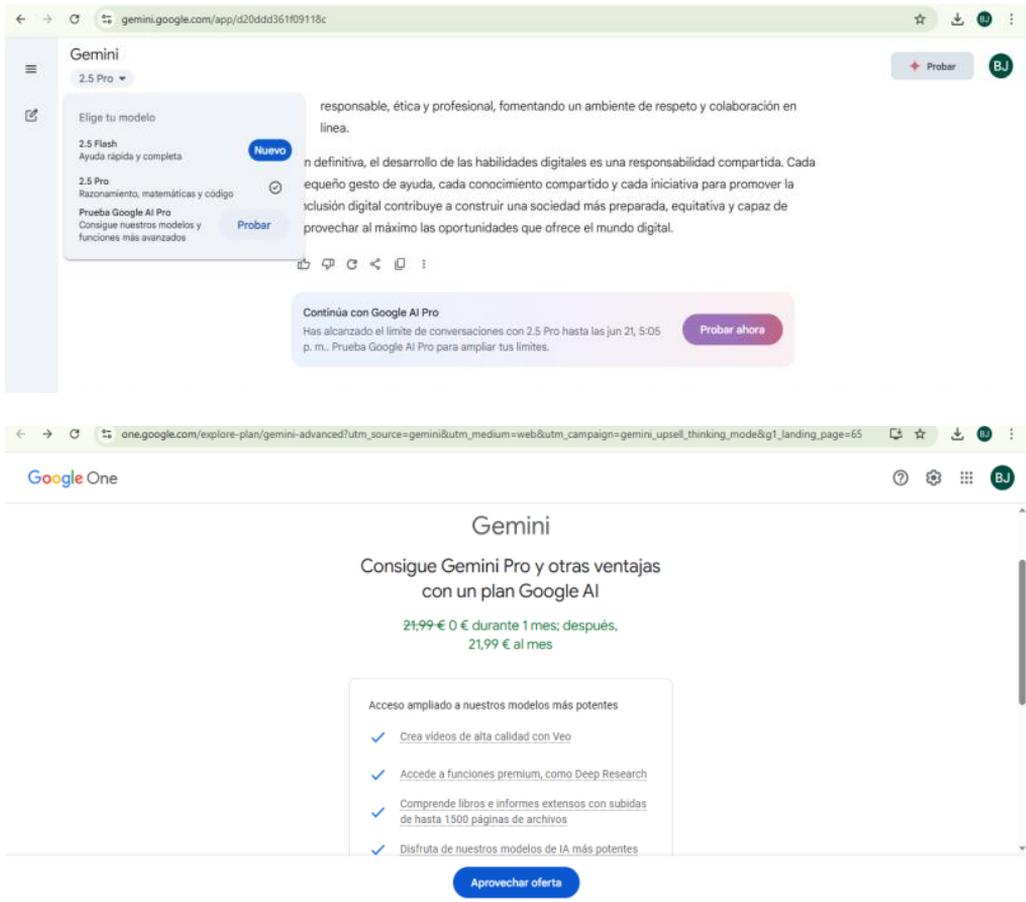


Figura 5. Fuente: Google (20/06/2025) Gemini (Modo prueba 2.5 Pro) [Herramienta de IA] Disponible en: <https://gemini.google/subscriptions/?hl=en-GB>

Del ejercicio realizado con el modelo Gemini Pro5, se puede analizar que las potencialidades que brindaba al inicio fueron reducidas debido a que al ser un modelo de negocio para la empresa Google, han dado a conocer al principio sus potencialidades sin ningún tipo de costo, brindando respuestas y creación de todo tipo de solicitudes con relación a distintas preguntas tanto de texto, imagen o video. Sin embargo, tal como la misma señala, actualmente que la Generación y Análisis de Contenido, Multimodalidad (Imágenes/Video), Límites de Uso y Velocidad, Integración con Otras Aplicaciones y Nuevas Funciones Experimentales a las que se tenía acceso, se trasladaron con mayor potencialidad a la suscripción de pago. Es decir, un costo de 21,99 Euros en forma mensual, lo que significa al usuario aproximadamente 264 Euros por año, en Europa.

Inteligencia Artificial: Análisis de su impacto en la brecha digital

En América Latina, el costo se lleva al tipo de cambio del país correspondiente, que en el caso boliviano equivale aproximadamente a Bs. 176.- al tipo de cambio oficial. Sin embargo, este valor cambia al tipo de cambio paralelo en el que le costaría a un usuario en Bolivia la suscripción en Bs. 324.- en forma mensual, por la situación económica actual. Misma que en otros países de la región también existe la diferencia con el tipo de cambio de divisa internacional.

Del mismo modo, se hizo el ejercicio con ChatGPT para verificar los beneficios de las versiones gratis (Free) o Pago, en la que respondió que en la versión Free trabaja con el modelo GPT-3.5. la velocidad de respuesta es razonablemente rápida, pero más lenta que GPT-4. Permite escribir texto, no se puede crear imágenes ni analizar archivos, siendo ideal para tareas básicas de escritura, respuestas generales, y ayuda con preguntas comunes. En tanto que la versión Plus que tiene un costo de (\$20 USD/mes) cuenta con mayores potencialidades como el modelo GPT-4, y la más avanzada llamada GPT-4o que es más rápido, más preciso, y con mejor comprensión del lenguaje y del contexto, se puede manejar imágenes, archivos, y entradas de preguntas avanzadas, ayuda en código, matemáticas, redacción compleja, análisis de archivos, chat de voz, memoria, etc.

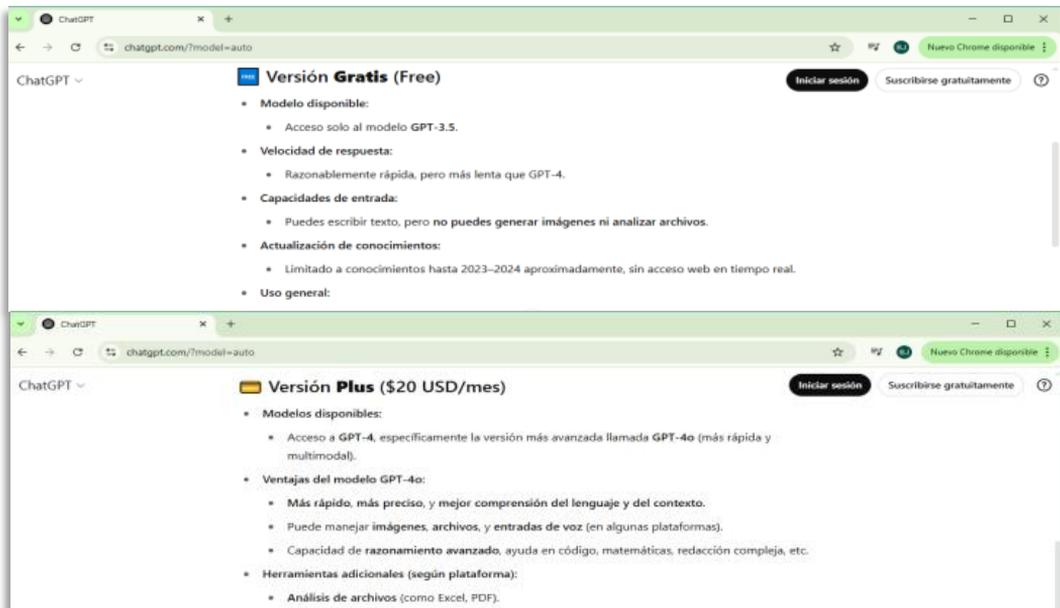


Figura 6. Fuente: OpenAI. (2025). ChatGPT (versión 3.5) [Herramienta de IA]. Disponible en: <https://chatgpt.com/?model=auto>

Entonces, se puede concluir de los dos ejercicios con los modelos Gemini y ChatGPT, que son herramientas que contribuyen a la brecha digital

económica, debido a que no todos tienen el poder adquisitivo para contar con todos los beneficios que brinda estos modelos de IA. De esa manera, persiste la desigualdad entre los usuarios de países con población económicamente pudiente en contra de aquellos que los recursos económicos bajos, que ciertamente podrán acceder a manejar la herramienta en la suscripción Free, pero con todas sus limitaciones. Por ejemplo, en Bolivia donde la situación económica, social y política tiene consecuencias en la sociedad, pensar en acceder a las bondades de estas plataformas queda fuera de lugar de la economía de las familias, el costo de los servicios básicos y sobre todo de la canasta familiar es su prioridad. De ese modo, sólo aquellos que tienen poder adquisitivo elevado podrán acceder a los modelos IA, en desmedro de aquellos cuyo salario es inferior o solo cuenta con el básico nacional, constituyéndose en una desventaja.

Con lo expresado hasta aquí, es posible finalizar indicando que existen dos tipos de usuarios, los que tienen habilidades digitales innatas y aquellas que requieren capacitación. Para los primeros, denominados nativos digitales, por su manejo intuitivo y no traumático de las TIC, quienes están a la vanguardia de la tecnología digital. Son responsables del mayor crecimiento de las plataformas digitales, tanto para efectuar consultas a la IA, como aquellas que les permiten comunicarse.

Tener acceso al Internet para este grupo es una necesidad y una forma de esparcimiento individualizado en la que invierten varias horas en sus computadoras o en sus celulares. Para los segundos, las brechas digitales también se constituyen un desafío, porque deben actualizarse y manejar los programas que aparecen, al mismo tiempo, contar con acceso al Internet hoy en día es parte de las actividades cotidianas, razón por la cual, debe ser una prioridad contar con habilidades digitales en tanto la tecnología vaya avanzando. De ahí, que las desigualdades o divisiones digitales estarán presentes de una u otra manera en la sociedad.

Con relación a la inteligencia artificial, estos modelos de respuesta tienen un impacto que ahonda la brecha digital económica, los precios no son accesibles para todos. Por tanto, las mayores potencialidades que puedan ofrecer estas empresas o corporaciones mundiales de generación de IA, continuarán generando un negocio que solamente los países con mayor poder adquisitivo tendrán acceso, en tanto, que los países y sociedades con menores ingresos estarán en desventaja, produciéndose así una fractura en lo concerniente al mundo digital.

7. Conclusiones

Las nuevas formas de trabajo y comunicación, han servido para establecer distintos tipos de brechas digitales, por las capacidades que existen entre

Inteligencia Artificial: Análisis de su impacto en la brecha digital

unos y otros, por el costo que significa acceder a la tecnología y saber utilizarla. En tanto existan usuarios que no cuenten con el suficiente dinero para pagar el costo de la tecnología ya sean plataformas, aplicaciones, redes sociales, programas e IA, la brecha digital no podrá cerrarse, y, al contrario, las grandes corporaciones que vayan generando nueva tecnología posiblemente dejen lo más básico a la suscripción gratuita, dejando lo más potente en las suscripciones de pago, lo que se constituye una brecha digital económica.

La brecha digital es un reflejo de las desigualdades y oportunidades existentes en Latinoamérica, aquellos países que están dando importancia al acceso a la banda ancha, tienen la oportunidad de desarrollar nuevas formas de trabajo, estudio y de alguna manera eso contribuye a una nueva fuerza laboral en la región.

Con una capacitación a gran escala en habilidades digitales se puede crear una economía más automatizada en distintas áreas como la contratación de personal, el acceso a créditos, la atención médica e incluso en el sistema de justicia. Aspectos que deben ser considerados al momento de buscar alternativas para incrementar el acceso a Internet y cerrar las brechas digitales.

El acceso a la infraestructura digital (internet de alta velocidad, dispositivos adecuados) sigue siendo un desafío en distintos países en las zonas rurales y comunidades marginadas de la región. Sin una conectividad asequible y de calidad, además de no contar con las habilidades digitales necesarias para interactuar con tecnologías complejas, una parte significativa de la población queda excluida de los beneficios de la tecnología y del uso de plataformas IA.

Una oportunidad que tienen los países y las sociedades latinoamericanas, es optar por buscar financiamiento o convenios internacionales con las empresas tecnológicas, para que los ciudadanos accedan a menor costo a las plataformas digitales y al manejo de las plataformas de inteligencia artificial, de ese modo, un Estado estará en posibilidad de crear ventajas competitivas para los ciudadanos a través de plataformas de aprendizaje, independientemente de su ubicación geográfica, crear nuevas formas de comunicación, nueva fuerza laboral digital, y el acceso a una inmensa cantidad de recursos digitales globales. Con estas proyecciones, las brechas digitales podrían cerrarse y los ciudadanos afianzar más su formación y conocimiento para la generación de mayores oportunidades de desarrollo económico y social.

Referencias

- Aladro, Eva** (2011) La Teoría de la Información ante las nuevas tecnologías de la comunicación CIC. Cuadernos de Información y Comunicación, vol. 16, pp. 83-93, Madrid, España. Disponible en: <http://www.redalyc.org/articulo.oa?id=93521629005>
- Autoridad de Regulación y Fiscalización de Telecomunicaciones y Transportes (ATT)** (2024) ESTADO DE SITUACIÓN ACTUAL DEL INTERNET EN BOLIVIA. Bolivia. Disponible en: https://www.att.gob.bo/sites/default/files/archivos_portada/2024-08/Estado%20de%20Situaci%C3%B3n%20Actual%20del%20Internet%20en%20Bolivia%202023.pdf
- Best Broadband Deals** (2024) Global broadband pricing league table 2024. Disponible en: <https://bestbroadbanddeals.co.uk/broadband/pricing/worldwide-comparison/>
- Best Broadband Deals** (2025) Who are you. Disponible en: <https://bestbroadbanddeals.co.uk/about/>
- Cambridge Dictionary** (2024) Artificial Intelligence. Disponible en: <https://dictionary.cambridge.org/es/diccionario/ingles/artificial-intelligence>
- Echazarreta, Carmen y Gürkan Hasan** (2022) Las brechas digitales de género. Colloquia Revista de Pensamiento y Cultura. Vol. 9, págs. 30 - 49. Disponible en: <https://doi.org/10.31207/colloquia.v9i0.133>
- García, Antonio, Iglesias, Enrique, Puig, Pau, y Dalio, Maribel** (2023). Informe anual del Índice de Desarrollo de la Banda Ancha: brecha digital en América Latina y el Caribe. Disponible en: <https://doi.org/10.18235/0004960>
- Google** (2025) Gemini AI. Disponible en: <https://gemini.google.com/>
- Hernández-Fuentes, Angie** (2022) Cooperación digital y soberanía tecnológica para cerrar la brecha digital en la cuarta revolución industrial. Oasis, núm. 36, 2022, Julio-diciembre, pp. 77-94. Colombia. Disponible en: <https://doi.org/10.18601/16577558.n36.06>
- Lazcano, Miguel** (26 de agosto de 2023). Más de la mitad de hogares en el país tiene acceso a internet. La Razón. Obtenido de <https://www.la-razon.com/economia/2023/08/26/mas-de-la-mitad-de-hogares-en-el-pais-tiene-acceso-a-internet/#:~:text=De%20acuerdo%20con%20la%20%C3%BAltima,con%20el%2056%2C89%25.>
- León, Cristian, Méndez, Lu y Machaca, Wilmer** (2022). No hay señal para la Educación. La Paz. Bolivia: Friedrich-Ebert-Stiftung en Bolivia (fes Bolivia). Disponible en: <https://library.fes.de/pdf-files/bueros/bolivien/19602.pdf>
- Lombana, Andrés** (2018) La evolución de las brechas digitales y el auge de la Inteligencia Artificial (IA). Revista Mexicana de Bachillerato a Distancia, número 20, año 10, México.

Inteligencia Artificial: Análisis de su impacto en la brecha digital

- Merriam-Webster** (2024) Inteligencia artificial. Disponible en: <https://www.merriam-webster.com/dictionary/artificial%20intelligence>
- Pizzi, Alejandro, Pecourt, Juan y Rius-Ulldemolins, Joaquim** (2023). De la "brecha digital" al control de internet: Usos, actitudes y participación digital en España. *Revista Española de Sociología*, 32(3), 1-20. Disponible en; <https://doi.org/10.22325/fes/res.2023.178>
- OpenAi** (2025) CharGPT Get started. Disponible en: <https://chatgpt.com/>
- Oracle** (2024). *Oracle OCI*. Disponible en: <https://www.oracle.com/mx/artificial-intelligence/machine-learning/what-is-machine-learning/>
- Real Academia Española (RAE)** (2024) Diccionario de la Real Academia Española. Disponible en: <https://dle.rae.es>
- Rodríguez, Mariola** (2023). Los intentos de regulación de la Inteligencia Artificial: Derecho Comparado entre España, Puerto Rico y Estados Unidos. *Derecom.* 35. Puerto Rico. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=9352231>
- Vargas, Marcia** (1996). La revolución de la información. FENIX. 17-33. Perú. Disponible en: https://www.researchgate.net/publication/348337135_La_revolucion_de_la_informacion

Equipo editorial, las autoras y los autores

Prof. Dr. Cástor David Mora

david.mora@uni-heidelberg.de

El Prof. Dr. Cástor David Mora ha sido docente e investigador en educación matemática, didáctica general, estadística, didácticas específicas, epistemología, planificación estratégica en educación, neurodidáctica, metodología de la investigación, paradigmas de investigación científica, ciencias interdisciplinarias, praxis innovadoras de aprendizaje y enseñanza, educación técnica, teoría de la actividad, etc. Ha desarrollado investigaciones y escrito/publicado artículos, libros y ensayos sobre temas de interés sociopolítico, educativo y científico. Igualmente, ha trabajado en diversos países e instituciones científicas nacionales e internacionales, tales como MPPE, UCV y UDO (Venezuela), UMSA, UNSXX y IICAB (Bolivia), UNAN (Nicaragua), UH, UB y HGGS-UH (Alemania). Fue docente de física, matemáticas y educación técnica en la República Bolivariana de Venezuela durante más de veinte años. Actualmente coordina el Programa de Cooperación Sur-Sur-Norte de la HGSS.



Dra. Astrid Wind

astrid.wind@hggs.uni-heidelberg.de



La Dra. Astrid Wind es la Directora General de la Escuela de Graduados en Humanidades y Ciencias Sociales de Heidelberg (HGGS). Antes de incorporarse a la Universidad de Heidelberg en 2013, obtuvo su Dr. Phil en la Universidad de Oxford y luego trabajó durante siete años como profesora e investigadora en La Paz (Bolivia), donde también dirigió el programa de doctorado de la Universidad Nacional „Siglo XX“ en Llallagua y el Instituto Internacional de Integración del Convenio Andrés Bello en La Paz. Acompañó la fundación de las universidades indígenas de Bolivia y publicó libros y artículos sobre las universidades indígenas

en distintas regiones del mundo. Durante los últimos años ha publicado, en el marco de la cooperación Sur-Sur-Norte de la HGSS, una variedad importante de artículos científicos interdisciplinarios sobre diversos temas de interés internacional.

Dra. Yanett Coromoto Segovia

yanett.segovia6@gmail.com

Profesora Titular en la Universidad de Los Andes (ULA) (Mérida, Venezuela) y Doctora en Antropología de América por la Universidad Complutense de Madrid. Exdirectora de la Escuela de Criminología (ULA) y actual Coordinadora del Doctorado en Ciencias Humanas de la ULA (desde 2022). Cofundadora del Doctorado en Estudios Territoriales en la Universidad de Caldas (Colombia). Su investigación se centra en epistemologías críticas, subjetividades y etnografías comprometidas, abordando antropología de las emociones, violencia, culturas de paz y migraciones. Destaca su trabajo con el pueblo indígena wayuu en la frontera colombo-venezolana, así como la codirección del proyecto „Escuela Convivencial para la Paz“. Es cofundadora de los grupos: „Vidas, Violencias y Con_Vivencias en América Latina y El Caribe“ (VALEC), y „Estudios Contemporáneos sobre Digitalidad e Historias Disruptivas“ (GECDDH). Miembro fundador de la: Red Iberoamericana de Etnografías Colaborativas y Comprometidas (ALA), Red Iberoamericana de Precariedades y Persistencias, y Red Latinoamericana de Biopolítica (CLACSO).



Daniel Quintero Rodríguez

acacio@ula.ve



Historiador y Abogado egresado de la Universidad de Los Andes (ULA) (Mérida, Venezuela), con estudios de Maestría en el Instituto de Altos Estudios Nacionales (IAEN) de la República del Ecuador, actualmente concluyendo sus estudios doctorales en Ciencias Humanas en ULA. Se desempeña como docente de la Facultad de Humanidades y Educación de la ULA, siendo responsable de las cátedras: „Historia de las Ideas Políticas“, „Historia del pensamiento Político en la Era Digital“ y el seminario „Pensamiento Político en la Sociedad de la Información“. Es cofundador y coordinador del Grupo de Estudios Contemporáneos sobre Digitalidad e Historias Disruptivas (GECDDH). Cuenta con más de quince años desarrollando investigaciones que involucran la imbricación tecnológica y social para conocer sus afectaciones en el ser humano. Ha sido ponente en eventos nacionales e internacionales, así como autor de artículos científicos en revistas de diversas partes del mundo. Asimismo, ha sido co/autor y editor de distintos libros. Entre sus líneas de investigación se encuentran la Digitalidad, el Capitalismo Cognitivo, la Etnografía de lo Intangible, los simulacros tecnológicos y las re-significaciones virtuales. Es miembro de la Red Iberoamericana de Investigación en Imaginarios y Representaciones (RIIR).

Vanessa Márquez Vargas

vamv@ula.ve

Docente e investigadora de la Facultad de Humanidades y Educación de la Universidad de Los Andes (ULA) (Mérida, Venezuela). Candidata a Doctor en Ciencias Humanas, magíster en Literatura Iberoamericana y Licenciada en Letras: Lengua y Literatura Hispanoamericana y Venezolana, por la Universidad de Los Andes. Sus áreas de interés académico y de investigación abarcan un amplio espectro de temáticas relacionadas con la cultura latinoamericana, literatura hispanoamericana y venezolana, ciencias del significado y educación. Tiene publicaciones y colaboraciones en Revistas, Centros e Institutos académicos nacionales e internacionales. Actualmente es becaria la Fundación Baden-Württemberg STIPENDIUM en el Programa de Intercambio Sur-Sur-Norte de la Escuela de Posgrado de Humanidades y Ciencias Sociales de Heidelberg (HGGS), Universidad de Heidelberg, Alemania durante el semestre de verano 2025.



Adrián Navia Segovia

adriansvia@gmail.com



Arquitecto egresado de la Universidad de Los Andes (ULA) (Mérida, Venezuela), con trayectoria en docencia universitaria, investigación y desarrollo de proyectos comunitarios. Actualmente se desempeña como Profesor de Taller de Diseño Arquitectónico en la Facultad de Arquitectura de la misma casa de estudios, cargo que ejerce desde 2018. Posee formación de posgrado en Filosofía y es doctorando en Ciencias Humanas, lo que enriquece su perspectiva crítica en torno a la arquitectura como disciplina social, cultural y política. Miembro activo de redes de investigación como VALEC, la Red Iberoamericana de Etnografías Colaborativas, y el Grupo de Estudios Contemporáneos sobre Digitalidad e Historias Disruptivas (GECDDH), mantiene una participación constante en espacios de reflexión y producción intelectual. Ha trabajado en proyectos como „Habitar el límite“, „Shabono Comunal“ y „El Spot“, enfocados en la recuperación de espacios públicos y el fortalecimiento de vínculos comunitarios. Además, ha sido tutor de trabajos de grado, ponente en congresos nacionales e internacionales, y autor de publicaciones académicas arbitradas. Ha recibido diversos reconocimientos, incluyendo el tercer lugar en el concurso „Ciudades Brillantes“ (2022) y distinciones por ponencias en encuentros académicos como el ELEA (2012).

Nathalie Miret González

nattymiret@gmail.com

Licenciada en Derecho (2017). Especialista en Negocios Internacionales (2018). Máster en Criminología (2024). Candidata Doctoral en Ciencias Jurídicas. Profesora Auxiliar de Derecho Penal y Criminología, y profesora principal de Práctica Jurídica con enfoque de derechos (Ejercicio Jurídico II), y Coordinadora de las Clínicas Jurídicas, Facultad de Derecho, Universidad de La Habana, Cuba. Ha impartido docencia en los colectivos de asignatura Derecho Penal General, Penal Especial, Procesal Penal, Criminología, Derecho Penal de Ejecución, Derecho Penal Económico y Derecho Penal Internacional. Profesora Principal de Año Académico en el Curso por Encuentro. Ha publicado artículos científicos en revistas especializadas indexadas y ha sido ponente en eventos científicos nacionales e internacionales. Investigadora y miembro de proyectos nacionales e internacionales. Miembro de las redes de investigación Red de Derecho América Latina y el Caribe (REDALC); Red Regional de Estudios sobre la violencia basada en género (RREVBG); Red de Juventudes y Red de apoyo a publicaciones en revistas académicas (RAPRE). Miembro del Capítulo Provincial de Derecho Penal de la Unión de Juristas de La Habana. Se ha desempeñado en el ejercicio de la profesión como Especialista en Negocios Internacionales, Jurista de Comercio Internacional y como Fiscal Provincial de La Habana, de la Fiscalía General de la República.



Alberto Carballo Soca

alberto.carballo@psico.uh.cu



Doctorando en Ciencias Psicológicas por la Universidad de La Habana, donde previamente obtuvo la Licenciatura en Psicología (2020) y el Máster en Psicología Educativa (2022). Posee además un Máster en Educación Virtual por la Universidad de las Ciencias Informáticas (2024). Ostenta las categorías de Profesor Auxiliar e Investigador Agregado. En la Facultad de Psicología de la Universidad de La Habana, ejerce como Jefe del Departamento de Formación Básica y es el Profesor Principal de las asignaturas de Metodología de la Investigación Psicológica I y II. Es también Miembro de la Junta Directiva del Grupo Provincial Habana de la Sociedad Cubana de Psicología.

Debby Avendaño Sánchez

debbyavendano@gmail.com

Arquitecta, egresada de la Universidad de Los Andes (ULA) (Mérida, Venezuela) y Magister Scientiarum en Historia de la Arquitectura y el Urbanismo en la Universidad Central de Venezuela. Docente de las cátedras sobre Historia de la Arquitectura Antigua, Moderna y Contemporánea, adscrita al Departamento de Materias Históricas y Humanísticas de la Facultad de Arquitectura y Diseño (ULA). Docente de la cátedra Crítica de la Arquitectura en la Maestría de Historia, Teoría y Crítica de la Facultad de Arquitectura y Diseño (ULA). Miembro del Grupo de Investigación sobre Patrimonio, Arquitectura y Ciudad (PAC), Facultad de Arquitectura y Diseño (ULA). Actualmente cursa un doctorado en Ciencias Humanas en la ULA y es becaria del Programa de Intercambio Sur-Sur-Norte de la Escuela de Posgrado de Humanidades y Ciencias Sociales de Heidelberg (HGGs), en la Universidad de Heidelberg (Alemania), en el semestre de verano 2025. Sus líneas de investigación incluyen: Historia de la Arquitectura Contemporánea, Historia de la Arquitectura Moderna en Venezuela, Historia de la Arquitectura Religiosa, Crítica de la Arquitectura y Patrimonio Arquitectónico.



Beatriz Chuquimia Aranda

mscbjca@gmail.com



Boliviana, nacida en la ciudad de La Paz, es Informática, Abogada y docente de posgrado. Profesional titulada por la Universidad Mayor de San Andrés en La Paz, Bolivia, asimismo, Magíster Scientiarum en Educación Superior del Centro Psicopedagógico en Educación Superior (CEPIES) de la Universidad Mayor de San Andrés y Magister en Derecho de los Servicios Financieros: Banca y Valores de la Universidad Andina Simón Bolívar. Cuenta con el Diplomado en Neuroeducación por la Universidad

Católica Boliviana San Pablo, Diplomado en Metodología de la Investigación Cualitativa y Diplomado en Organización y Administración Pedagógica del Aula en Educación Superior por la Universidad Mayor de San Andrés. Ha sido consultora en el área de informática y programación, docente universitaria en distintas asignaturas en informática, derecho y matemática, ejerció el cargo de Rectora del tecnológico Santo Toribio de Mogrovejo (2008-2010), es docente de posgrado en distintos programas posgraduales en Educación Superior, Informática y Matemática desde el año 2009 hasta la actualidad. Se desempeña como profesional independiente y actualmente es doctoranda en Ciencias de la Informática e Ingeniería de la Universidad Nacional „Siglo XX“ en Bolivia que cuenta con la cooperación de la HGGs (Universidad de Heidelberg).