INTELIGENCIA ARTIFICIAL

EXPERIENCIAS Y REFLEXIONES SOBRE LA INVESTIGACIÓN CIENTÍFICA

ALEXANDRO ESCUDERO-NAHÓN
EDITOR

Transdigital®
editorial

INTELIGENCIA ARTIFICIAL

EXPERIENCIAS Y REFLEXIONES SOBRE LA INVESTIGACIÓN CIENTÍFICA

ALEXANDRO ESCUDERO-NAHÓN

Editor

Alejandro Guadalupe Rincón Castillo, Alexandro Escudero-Nahón, Alma Delia Otero Escobar, Andrea Sánchez-Ruiz, Andrés Valencia Sánchez, Antonio Franco Vadillo, Antonio Juan Capistran Abundez, Arturo González Torres, Aura Patricia Hernández Olicón, Blanca Cecilia López Ramírez, Cándida Marcela Rodríguez Chávez, Carlos Enrique Levet Rivera, Carlos Zepeda-Lugo, Carolina Medina García, Cecilia Esperanza Ostos Cruz, Cesaire Chiatchoua, Christian Paulina Mendoza Torres, Clara Rosalva Mercado-López, Claudia Margarita García Paulin, Daniel Alberto Mejía Herrera, Diego Escudero-Sánchez, Elena Elsa Bricio-Barrios, Elizabeth Vanessa Teniente Gasca, Elsa Suárez Jasso, Emma Patricia Mercado-López, Erik Carbajal-Degante, Fabián Gómez Santiago, Francisco Antonio Torres-Espriú, Francisco de Jesús Mata Gómez, Francisco Guadalupe Avendaño Esparza, Georgina del Carmen Mota Valtierra, Gerardo Quiroz Bojorges, Guillermo Vargas Rodríguez, Héctor Alfredo Baptista Gonzales, Humberto Aguirre Becerra, Inés Guadalupe Germán Aguilar, Itzia Nallely Guzmán Mejía, , Ivette Selene Marañón Lizárraga, José Antonio Cisneros Jiménez, José Cristóbal Solís Pollorena, José Luis Bautista López, Juan Carlos Lobato-Valdespino, Julia Dolores Toscano Garibay, Karina Guadalupe Cortina Calderón, Leonardo Eliphas Daza Ramírez, Leonardo Ledesma Domínguez, Lucia Morales-Morales, Luis Alonso Castañeda Negrete, Luis Javier Raúl Obregón Herrin, Luis Ramón Carreño Durán, Luz Angélica Mondragón del Angel, Ma. Cristina Vázquez Hernández, Manuel Ramón González Herrera, Marcos Sanchez-Lizarraga, Mariajosé López Laiza, Mario Alberto Domínguez-Rovira, Marysol Estrella Hernández García, Miguel Ángel Medina Romero, Mireille Toledo Blas, Modesta Lorena Hernández Sánchez, Mónica Lorena Sánchez Limón, Nallely Guadalupe Hernández Hernández, Octavio Reyes López, Pavel David Ulises Avendaño López, Ramar Mendoza Díaz, Rita Ávila Romero, Rodrigo Ochoa Figueroa, Salvador Ortiz Santos, Santiago Arceo-Diaz, Tania Haidée Torres Chávez, Tomás Peralta Palazón, Vitervo López-Caballero y Xóchitl Trujillo-Trujillo.

Autores y autoras

Transdigital®
editorial

Título original: Inteligencia artificial: experiencias y reflexiones sobre la investigación científica / Alexandro Escudero-Nahón (Editor) — Ciudad de Querétaro, México: Editorial Transdigital, 2026 — 457 páginas.

International Standard Book Number (ISBN): 978-968-9724-25-4.

Digital Object Identifier (DOI) del libro: https://doi.org/10.56162/transdigitalbc12

Clasificación DEWEY. Materia: 370.7–Estudio y enseñanza de la educación. Tipo de Contenido: Libros universitarios. Clasificación thema: JN–Educación. Tipo de soporte: libro digital gratuito descargable. Formato: PDF. Tamaño: 8.3 Mb.

Este libro es una publicación de acceso abierto con los principios de Creative Commons Atribution 4.0 International License (CC BY-NC-SA). Esta licencia permite a los reutilizadores distribuir, remezclar, adaptar y desarrollar el material en cualquier medio o formato únicamente con fines no comerciales y siempre que se otorgue la atribución al creador. Si remezcla, adapta o construye sobre el material, debe licenciar el material modificado bajo términos idénticos.

Esta obra ha sido dictaminada por pares académicos expertos con el método de doble ciego. Los dictámenes están resguardados en los archivos de la Editorial *Transdigital*.

D.R. 2026 Alexandro Escudero-Nahón (Editor).

D.R. 2026 Autores y autoras.

D.R. 2026 Sello Editorial *Transdigital*.

Sociedad de Investigación sobre Estudios Digitales, S. C. Nombre de marca: *Transdigital*. Dirección: Circuito Altos Juriquilla 1132. Colonia Altos Juriquilla. C. P. 76230, Juriquilla, Querétaro, México.

+52 (442)301 32 38. editorial@transdigital.mx www.editorial.transdigital.mx

Registro en el Padrón Nacional de Editores como agente editor Sociedad de Investigación sobre Estudios Digitales, S. C., con el Dígito Identificador 978-607-99594.

Afiliación a la Cámara Nacional de la Industria Editorial Mexicana (CANIEM) con el número 4069, de conformidad con el artículo 17 de la Ley de Cámaras Empresariales y sus Confederaciones en vigor.

Registro Nacional de Instituciones y Empresas Científicas y Tecnológicas de la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) de México con el folio: RENIECYT 2400068.

Sugerencia de referencia para el libro en APA 7a. edición:

Escudero-Nahón, A. (2026) (Editor). *Inteligencia artificial: experiencias y reflexiones sobre la investigación científica.* Editorial Transdigital. https://doi.org/10.56162/transdigitalbc12

Contenido

00.
Análisis de las experiencias y reflexiones sobre la inteligencia artificial en la investigación científica 9
Alexandro Escudero-Nahón y Diego Escudero-Sánchez

01.
Condiciones sociales en la planeación estratégica para la adquisición de inteligencia artificial en las organizaciones 29
José Antonio Cisneros Jiménez

02.
Employment Situation for Recent University Graduates in Mexico City (2020-2024) 43
Mariajosé López Laiza, Rita Ávila Romero y Cesaire Chiatchoua

03.
Inteligencia artificial y paleogenómica predictiva: Incompatibilidad Rh y Kell en el colapso demográfico neandertal 58
Luis Ramón Carreño Durán, Aura Patricia Hernández Olicón, Antonio Franco Vadillo, Mireille Toledo Blas, Fabián Gómez Santiago y Héctor Alfredo Baptista Gonzales

04.
Justicia algorítmica y gobernanza ética ante los sesgos de la inteligencia artificial 77
Alma Delia Otero Escobar, Cecilia Esperanza Ostos Cruz y Elsa Suárez Jasso

05.
Inteligencia artificial generativa y ética profesional en la contaduría pública 96
Leonardo Eliphas Daza Ramírez y Francisco de Jesús Mata Gómez

06.
Integración de la inteligencia artificial en procesos de investigación en instituciones educativas en Zacatecas, México 123
Alejandro Guadalupe Rincón Castillo, Cándida Marcela Rodríguez Chávez, Luis Alonso Castañeda Negrete y Daniel Alberto Mejía Herrera

07.
¿Pueden las máquinas salvar al maíz? Aprendizaje profundo para la detección de plagas 142
Antonio Juan Capistran-Abundez, Vitervo López-Caballero, Lucia Morales-Morales y Andrea Sánchez-Ruiz

08.
Triple convergencia en la era de la transformación digital del turismo: ciencia de datos, inteligencia analítica y gestión de destinos 159
Manuel Ramón González Herrera y Carolina Medina García

09.
Uso de la inteligencia artificial en el coaching empresarial (2024): revisión de literatura 175
Arturo González Torres, Gerardo Quiroz Bojorges y Pavel David Ulises Avendaño López

10.
El uso de la inteligencia artificial en los procesos de investigación científica 193
Marysol Estrella Hernández García

11.
Hacia una nueva praxis de ciencia abierta dominada por datos masivos e inteligencia artificial generativa 208
Erik Carbajal-Degante y Leonardo Ledesma-Domínguez

12.
La inteligencia artificial en el derecho: uso práctico y transición regulatoria 224
Carlos Enrique Levet Rivera, Modesta Lorena Hernández Sánchez y Ramar Mendoza Díaz

13.
Envejecimiento digno en México: detección en tiempo real del nivel de riesgo de sarcopenia mediante inteligencia artificial 238
Santiago Arceo-Diaz, Xóchitl Trujillo-Trujillo y Elena Elsa Bricio-Barrios

14.
El impacto de la inteligencia artificial generativa en la arquitectura editorial y el crecimiento exponencial de la producción científica en salud 251
Salvador Ortiz Santos, Georgina del Carmen Mota Valtierra, Humberto Aguirre Becerra, Blanca Cecilia López Ramírez y Ma. Cristina Vázquez Hernández

15.
Factores críticos de éxito para implementar la inteligencia artificial en las unidades de cuidados intensivos 265
Carlos Zepeda-Lugo, Marcos Sanchez-Lizarraga e Ivette Selene Marañón Lizárraga

16.
Escucha humana e inteligencia artificial:
límites y alcances en la investigación musical 280
José Luis Bautista López, Guillermo Vargas Rodríguez y
Luis Javier Raúl Obregón Herrin

17.
Evolución de la representación gráfica: del trazo humano al algoritmo..... 295
Luz Angélica Mondragón del Angel e Inés Guadalupe Germán Aguilar

18.
Conocimiento o apariencia:
el estatuto epistémico de la inteligencia artificial generativa..... 312
Tomás Peralta Palazón

19.
Entre el prompt y el diseño: experiencias de co-creación
crítica humano-inteligencia artificial en educación superior..... 325
Juan Carlos Lobato-Valdespino y Claudia Margarita García Paulín

20.
Aplicación de la inteligencia artificial en la gestión administrativa
de las micro, pequeñas y medianas empresas del sector de la construcción 343
Elizabeth Vanessa Teniente Gasca, Octavio Reyes López y
Christian Paulina Mendoza Torres

21.
Modelos y aplicaciones de machine
learning en la estratificación de riesgo clínico..... 359
Julia Dolores Toscano Garibay

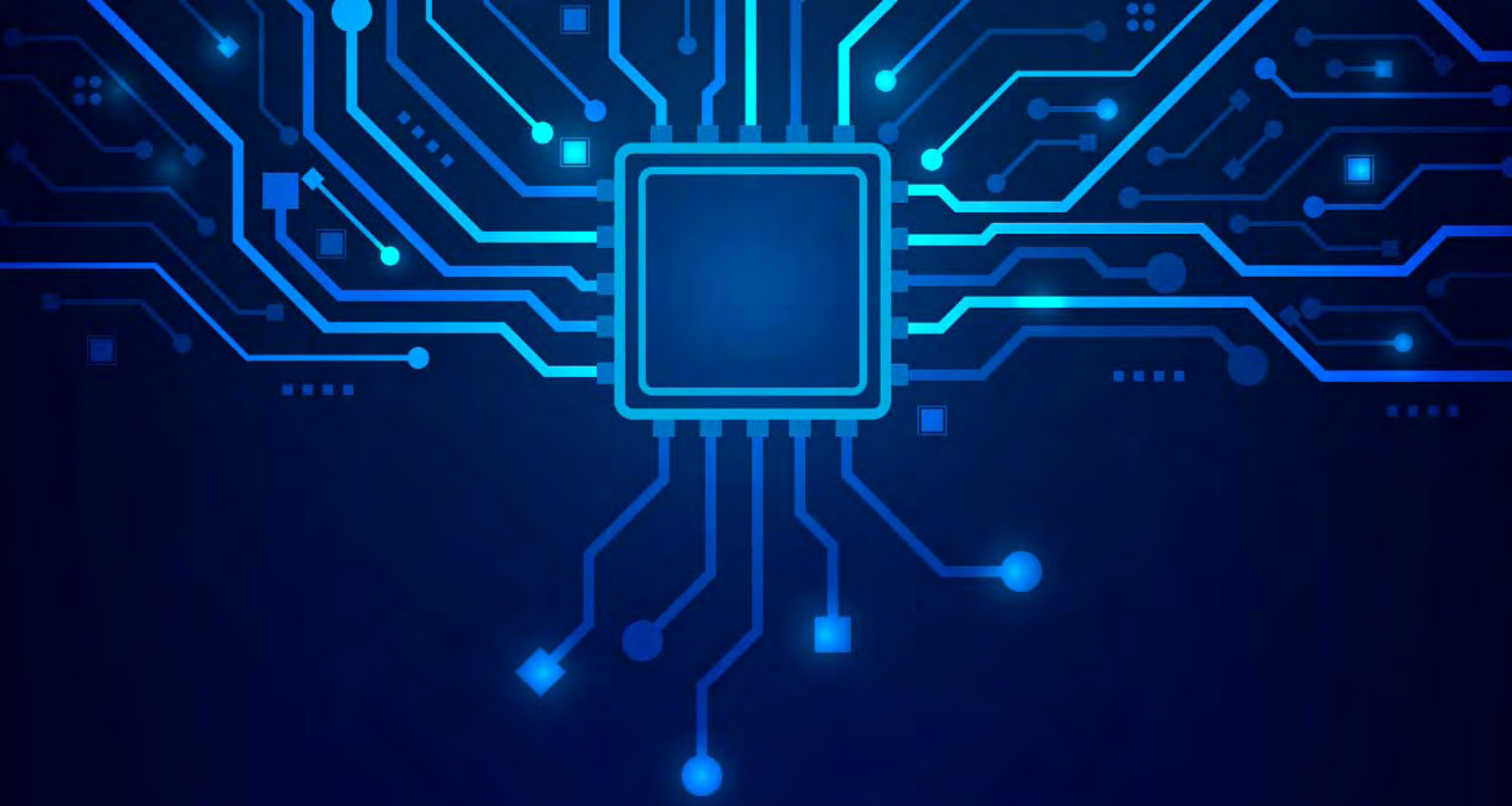
22.
Entre principios y práctica: revisión de marcos regulatorios
y éticos sobre inteligencia artificial en la investigación científica 374
Miguel Ángel Medina Romero, Tania Haidée Torres Chávez y Rodrigo Ochoa Figueroa

23.
Inteligencia artificial en el proceso de
investigación científica: usos, experiencias y limitaciones..... 389
Emma Patricia Mercado-López y Clara Rosalva Mercado-López

24.
La mediación epistémica de la inteligencia artificial generativa en la ciencia contemporánea 406
Andrés Valencia Sánchez y José Cristóbal Solís Pollorena

25.
Análisis documental de la adopción y humanización de asistentes digitales basados en inteligencia artificial 422
Karina Guadalupe Cortina Calderón, Nallely Guadalupe Hernández Hernández y Mónica Lorena Sánchez Limón

26.
Del andamiaje con inteligencia artificial a la autonomía: evaluación de código estadístico en ingeniería 442
Francisco Antonio Torres-Espriú, Itzia Nallely Guzmán Mejía, Francisco Guadalupe Avendaño Esparza y Mario Alberto Domínguez-Rovira



00.

Análisis de las experiencias y reflexiones sobre la inteligencia artificial en la investigación científica

Alexandro Escudero-Nahón
Universidad Autónoma de Querétaro, México
ORCID: 0000-0001-8245-0838

Diego Escudero-Sánchez
Universidad Autónoma de Querétaro, México
ORCID: 0000-0001-6128-886X

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.00

00.

Análisis de las experiencias y reflexiones sobre la inteligencia artificial en la investigación científica

Introducción

El libro *Inteligencia artificial: experiencias y reflexiones sobre la investigación científica* constituye una obra colectiva que reúne capítulos desarrollados por autoras y autores provenientes de disciplinas, instituciones y contextos muy diversos. En conjunto, estos trabajos ofrecen un panorama complejo de la forma en que la inteligencia artificial (IA) está reconfigurando la investigación científica, no sólo como repertorio de herramientas, sino como fenómeno socio-técnico que atraviesa estructuras organizacionales, marcos regulatorios, culturas profesionales y concepciones de conocimiento.

La obra se organiza en torno a experiencias situadas de uso de IA en campos tan distintos como la planeación estratégica organizacional, el análisis del mercado laboral de egresados universitarios, la paleogenómica, la justicia algorítmica, la ética profesional en la contaduría pública, la investigación educativa, el turismo, la agricultura, la salud, el derecho, la música, el diseño gráfico y la propia arquitectura editorial de la ciencia en salud. Esta pluralidad de escenarios no es un rasgo anecdótico, sino uno de los aportes centrales del libro: obliga a abandonar cualquier perspectiva simplificadora que conciba la IA como una tecnología homogénea, universal y neutra, y la muestra, en cambio, como un entramado de prácticas, discursos y dispositivos cuyo significado se reconstruye en cada contexto de uso.

Desde una mirada comprensiva, puede señalarse que las contribuciones reunidas en el volumen se agrupan en tres grandes tipos de aportes. En primer lugar, se encuentran estudios empíricos de corte aplicado, en los que la IA se utiliza para abordar problemas específicos de investigación: la reconstrucción de fenotipos sanguíneos en poblaciones neandertales, la detección temprana de plagas en cultivos de maíz, la estratificación de riesgo clínico, la identificación de sarcopenia o la gestión administrativa en micro, pequeñas y medianas empresas.

En segundo lugar, el libro incluye trabajos que analizan procesos de adopción e integración de IA en organizaciones y entornos profesionales, tales como instituciones educativas, despachos de contaduría, equipos de coaching y unidades de cuidados intensivos. Finalmente, la obra ofrece capítulos de corte conceptual, normativo y filosófico que discuten la justicia algorítmica, los marcos regulatorios y éticos, la ciencia abierta y el estatuto epistémico de la IA generativa, así como experiencias de co-creación estético-investigadora entre humanos y sistemas algorítmicos.

El propósito de este capítulo inicial es realizar un análisis crítico de conjunto que permita articular dichas experiencias y reflexiones en una narrativa estructurada. Más que ofrecer una reseña capítulo por capítulo, se propone una lectura transversal organizada en ejes analíticos que cruzan la totalidad del libro. Estos ejes son: a) las condiciones sociales y organizacionales de adopción de tecnologías de IA; b) la IA como instrumento científico en distintos campos de conocimiento; c) las transformaciones del trabajo intelectual y de la práctica investigadora; d) los impactos en la comunicación científica, la ciencia abierta y la arquitectura editorial; e) las dimensiones éticas, jurídicas y de justicia algorítmica; y f) las implicaciones epistemológicas y culturales de la co-creación humano-máquina.

Desde esta perspectiva, la IA deja de ser un objeto unívoco y aparece como una constelación de procesos en los que intervienen actores humanos y no humanos, infraestructuras técnicas, normas formales e informales, valores, intereses y visiones de futuro. El análisis que aquí se presenta pretende, por un lado, ofrecer al lector una cartografía de la obra que facilite la lectura de los capítulos posteriores y, por otro, plantear problemas y preguntas que sirvan como marco para debates ulteriores en torno a la relación entre IA e investigación científica en contextos latinoamericanos y globales.

Condiciones sociales y organizacionales para la adopción de la inteligencia artificial

El capítulo que abre el libro, dedicado a las condiciones sociales en la planeación estratégica para la adquisición de IA en las organizaciones, establece un marco conceptual que resulta especialmente útil para comprender el resto de la obra. En él se sostiene que la incorporación de tecnologías de IA no puede entenderse como un acto puramente técnico de compra e instalación de sistemas, sino como un proceso complejo de transferencia tecnológica (TT) que abarca tres momentos articulados: adquisición, adaptación y asimilación.

Ambigüedad conceptual, discursos y problematización crítica

El capítulo inicia su argumentación llamando la atención sobre la ambigüedad y vaguedad del propio concepto de IAI. A través de un análisis comparado de definiciones provenientes de la informática, la ingeniería, las ciencias del trabajo y organismos internacionales, se muestra cómo la expresión *IA* designa, en distintos contextos, tanto un conjunto de técnicas y algoritmos como productos, procesos, sistemas autónomos y, en ocasiones, incluso imaginarios culturales sobre máquinas *inteligentes*.

Esta polisemia no se considera un mero problema terminológico, sino una condición que incide directamente en la manera en que las organizaciones perciben y gestionan la IA. Cuando un mismo término refiere, simultáneamente, a soluciones de analítica de datos, sistemas de recomendación, plataformas de automatización de procesos o interfaces conversacionales, se dificulta la tarea de evaluar riesgos y oportunidades y, sobre todo, de trazar estrategias de adopción alineadas con capacidades y necesidades concretas. El capítulo plantea, por ello, la necesidad de un ejercicio de *problematización* en sentido fuerte, entendido como un trabajo de pensamiento crítico orientado a desbordar la mera interrogación superficial para interrogar las condiciones de posibilidad de la IA como fenómeno social.

Tomando como referencia aportes de la sociología de Niklas Luhmann y de la teoría crítica, el autor propone sustituir una visión de la sociedad como conjunto de *sectores* por una noción de campos problemáticos, donde lo relevante no es sólo quiénes son los actores, sino qué problemas se formulan, desde qué posiciones y con qué tensiones. En ese marco, la IA aparece como un campo problemático en el que convergen discursos dominantes, críticas reactivas y críticas activas, que se despliegan en tres perspectivas: política-regulatoria, técnica-tecnológica y social-cultural.

Marco dialéctico y transferencia tecnológica

El capítulo introduce un esquema dialéctico que organiza estas perspectivas en torno a pares de conceptos en tensión. En la dimensión política-regulatoria, se contrasta, por ejemplo, la globalización de las tecnologías de información y comunicación (TIC) y la agenda neo-republicana con procesos de desregulación de la IA y de institucionalización polisémica, en los que las normas se formulan de manera fragmentaria y muchas veces responden más a lógicas de mercado que a consideraciones de justicia social. En la dimensión técnica-tecnológica, se discute el determinismo tecnológico frente a enfoques que subrayan la agencia

material y la contingencia de los desarrollos, así como debates sobre posthumanismo, transhumanismo y los límites técnicos de replicar la inteligencia humana. Finalmente, en la dimensión social-cultural, se examina la convergencia entre pensamiento humanista e imaginarios escatológicos, la centralidad de la información en la sociedad contemporánea y las condiciones sociales de aceptación o resistencia ante la IA.

Sobre esta base, se construye un marco analítico para pensar la estrategia de TT en organizaciones que desean adquirir soluciones de IA. La propuesta articula, en una tabla, distintas preguntas clave para cada fase del proceso (adquisición, adaptación y asimilación) y las vincula con dimensiones del entorno social (regulación tecnológica, vinculación institucional, madurez tecnológica, ética organizacional, fuerzas de mercado, funcionalismo técnico, eficiencia operativa, ergonomía y riesgos, dependencia técnica, accesibilidad, dinámicas de poder, valores culturales, liderazgo, difusión pública).

Este marco se aleja de una visión lineal de la innovación y subraya que, por ejemplo, una organización puede adquirir una tecnología con alto nivel de madurez técnica, pero fracasar en su adaptación si no cuenta con infraestructura compatible, capacidades de mantenimiento o mecanismos de formación del personal. Asimismo, puede asimilar funcionalmente un sistema (en términos de uso cotidiano), pero generar dinámicas de dependencia y pérdida de capacidades críticas si no se atienden los riesgos psicológicos y sociales asociados.

Desde una perspectiva crítica, el capítulo cuestiona la idea de que la IA sea una *oportunidad* que las organizaciones deben aprovechar cuanto antes para no quedar rezagadas. En lugar de ello, insiste en la necesidad de construir estrategias de adopción que reconozcan la contingencia del entorno, la multiplicidad de fuerzas sociales en juego y el carácter necesariamente experimental y revisable de las decisiones tecnológicas. Esta mirada permite leer el resto de capítulos del libro como exploraciones situadas de distintas combinaciones de adquisición, adaptación y asimilación de IA en contextos específicos.

1.3. Mercado laboral de egresados y brechas formativas

El segundo capítulo se centra en la situación de empleo de recién egresados universitarios en la Ciudad de México entre 2020 y 2024. Aunque no aborda la IA de forma directa en términos de uso de herramientas concretas, ofrece información relevante sobre el contexto en el que se insertan muchos de los sujetos que habrán de interactuar con tecnologías de IA en su práctica profesional e investigadora.

A partir de datos cuantitativos, el texto caracteriza una situación marcada por tasas de desempleo y subempleo apreciables, por la predominancia de contratos precarios y por una brecha significativa entre las competencias desarrolladas en la educación superior y las demandas del mercado laboral, que cada vez valora más habilidades digitales, flexibilidad y capacidades de análisis de información compleja. En este escenario, la irrupción de la IA y de otras tecnologías emergentes se convierte en un factor adicional que puede agravar desigualdades existentes si no se acompaña de políticas formativas y laborales inclusivas.

Desde la óptica de este capítulo de apertura, estos datos pueden leerse como una advertencia: la potencial adopción de IA en investigación y en sectores productivos no se distribuye sobre un terreno neutro, sino sobre estructuras laborales atravesadas por precariedad y segmentación. El riesgo de que la IA contribuya a generar *élite* de trabajadores hipercualificados capaces de diseñar, auditar y gobernar sistemas, y una masa de profesionales que sólo pueden operar interfaces opacas sin capacidad de agencia crítica, es una preocupación transversal que reaparece en otros capítulos.

Inteligencia artificial como instrumento científico en campos específicos

Un segundo eje analítico de la obra lo constituyen los capítulos que documentan usos de IA para abordar problemas científicos delimitados. Estos trabajos permiten observar de qué manera la IA se integra en diseños metodológicos concretos, qué tipo de preguntas facilita formular y qué limitaciones y dependencias genera.

Paleogenómica predictiva y colapso demográfico neandertal

El tercer capítulo, dedicado a la relación entre IA, paleogenómica e incompatibilidad Rh y Kell en el colapso demográfico neandertal, constituye un ejemplo paradigmático de integración de herramientas de aprendizaje automático en un programa de investigación que combina genética, paleoantropología y medicina transfusional. A partir de genomas de alta cobertura de individuos neandertales y denisovanos, el estudio utiliza modelos predictivos para inferir fenotipos sanguíneos y explorar las consecuencias que determinadas incompatibilidades inmunohematológicas pudieron tener en escenarios de hibridación entre poblaciones.

La IA desempeña, en este caso, un papel decisivo en dos niveles. En primer lugar, actúa como mecanismo de integración de grandes volúmenes de datos genómicos, permitiendo identificar patrones complejos que serían difíciles de detectar mediante procedimientos estadísticos convencionales. En segundo lugar, posibilita la construcción de modelos contrafactuales sobre la dinámica de poblaciones, al simular distintos escenarios de compatibilidad e incompatibilidad que ayudan a evaluar hipótesis sobre las causas del declive neandertal.

El capítulo ilustra un uso *epistémicamente productivo* de la IA: lejos de limitarse a automatizar tareas rutinarias, amplía el espacio de interrogación científica al hacer manejables combinaciones de variables y escenarios que exceden la capacidad humana de cálculo. No obstante, el propio diseño del estudio hace visibles las limitaciones de esta aproximación. Los modelos dependen de la calidad, diversidad y representatividad de las muestras genómicas disponibles, que a su vez están condicionadas por decisiones históricas de excavación, preservación y secuenciación. Así, la capacidad predictiva de la IA no elimina, sino que desplaza, las fuentes de incertidumbre y sesgo.

Desde una perspectiva crítica más amplia, la contribución de este capítulo al conjunto del libro consiste en mostrar que la IA puede contribuir a reescribir narrativas sobre procesos evolutivos, pero también recuerda que cualquier innovación metodológica debe evaluarse a la luz de los supuestos y limitaciones que arrastra el archivo de datos con el que opera.

Inteligencia artificial en salud: riesgo clínico, cuidados intensivos y envejecimiento

La obra incluye varios capítulos centrados en el uso de IA en el ámbito de la salud, en particular en la estratificación de riesgo clínico, en la implementación de sistemas de IA en unidades de cuidados intensivos y en la detección en tiempo real del riesgo de sarcopenia en personas mayores. Estos textos comparten el interés por explorar cómo modelos de *machine learning* pueden mejorar la capacidad de anticipar desenlaces adversos y, por ende, apoyar decisiones clínicas en entornos donde el tiempo y los recursos son limitados.

En el caso de los modelos de estratificación de riesgo, se describen procedimientos de construcción y validación de algoritmos entrenados con bases de datos clínicas heterogéneas, que integran variables sociodemográficas, biomarcadores, antecedentes y resultados previos. La IA se presenta como un instrumento capaz de identificar combinaciones de variables no triviales y de producir puntajes de riesgo que superan, en algunos indicadores, la capacidad de reglas clínicas tradicionales.

El capítulo dedicado a los factores críticos de éxito para implementar IA en unidades de cuidados intensivos introduce, además, una reflexión organizacional imprescindible. No basta con que un modelo prediga con precisión determinada probabilidad de deterioro o mortalidad; es necesario que la organización cuente con infraestructura tecnológica adecuada, con flujos de datos confiables, con protocolos de uso integrados a los procesos clínicos, con personal capacitado para interpretar y discutir las recomendaciones algorítmicas, y con marcos éticos que delimiten responsabilidades.

De manera complementaria, el capítulo sobre detección de sarcopenia mediante IA muestra cómo dispositivos y algoritmos pueden monitorear, casi en tiempo real, variables asociadas a masa y fuerza muscular para generar alertas que faciliten intervenciones preventivas. Este tipo de soluciones refleja el tránsito hacia una medicina cada vez más basada en datos continuos, en la que la frontera entre investigación y práctica clínica se vuelve difusa.

Desde un enfoque crítico, los capítulos coinciden en señalar riesgos y dilemas. Entre ellos destacan: la opacidad de modelos complejos que dificultan la explicación de por qué se asigna determinado nivel de riesgo a un paciente; la posibilidad de que los modelos reproduzcan sesgos presentes en datos históricos (por ejemplo, discriminaciones por género, etnia o condición socioeconómica); y el peligro de que la carga de trabajo y la presión institucional induzcan a una dependencia excesiva de las recomendaciones algorítmicas en detrimento del juicio clínico. En suma, la IA aparece como un recurso poderoso, pero también como un vector de posibles vulneraciones a la autonomía profesional y a la equidad en salud si se implementa sin salvaguardas.

Inteligencia artificial para sistemas agroalimentarios y medio ambiente

Otro ámbito en el que la IA se presenta como instrumento científico es el de los sistemas agroalimentarios, en particular en el capítulo titulado *¿Pueden las máquinas salvar al maíz? Deep learning en la detección de plagas*. En este trabajo se desarrolla y prueba un sistema basado en redes neuronales profundas capaz de identificar, a partir de imágenes, la presencia temprana de plagas en cultivos de maíz.

Metodológicamente, el capítulo describe la construcción de conjuntos de datos de imágenes, el entrenamiento del modelo, la evaluación de su precisión y la implementación de un prototipo para su uso en campo. La IA permite reconocer patrones visuales que, en ocasiones, resultan imperceptibles para el ojo humano en etapas iniciales de infestación,

lo cual abre la posibilidad de intervenciones más oportunas y de una reducción en el uso de agroquímicos.

Sin embargo, el planteamiento mismo del capítulo, condensado en la pregunta invita a una lectura más amplia. El texto subraya que la supervivencia de variedades nativas de maíz y la sostenibilidad de sistemas productivos campesinos dependen de factores económicos, políticos y culturales que exceden con mucho las capacidades de cualquier sistema algorítmico. El acceso a tecnologías de IA, la conectividad en zonas rurales, la disponibilidad de dispositivos adecuados, la pertinencia cultural de las soluciones propuestas y las formas de gobernanza de datos agrícolas son elementos que condicionan la efectividad y la justicia de tales innovaciones.

Este capítulo permite, por tanto, problematizar la relación entre IA, soberanía alimentaria y justicia ambiental. La IA puede convertirse en un aliado para el monitoreo y el manejo sostenible de cultivos, pero también corre el riesgo de consolidar la dependencia de pequeños productores respecto a plataformas tecnológicas controladas por grandes corporaciones si no se diseñan modelos de desarrollo tecnológico inclusivos y participativos.

Transformaciones del trabajo intelectual y la práctica investigadora

Un tercer eje que recorre el libro se refiere a los cambios que la IA introduce en el trabajo intelectual de quienes investigan, enseñan, asesoran organizaciones o ejercen profesiones basadas en el conocimiento. Aquí la IA no se presenta solamente como un recurso instrumental, sino como un agente que reorganiza la distribución de tareas, redefine qué se considera trabajo *propiamente humano* y obliga a revisar marcos deontológicos consolidados.

En este grupo se ubican, principalmente, los capítulos sobre ética profesional en la contaduría pública, integración de IA en procesos de investigación en instituciones educativas de Zacatecas, México, uso de IA en *coaching* empresarial y aplicación de IA en la gestión administrativa de micro, pequeñas y medianas empresas del sector construcción. Aunque pertenecen a campos distintos, todos comparten la preocupación por el modo en que la IA reconfigura la relación entre automatización, juicio experto y responsabilidad.

Integración de inteligencia artificial en la investigación educativa

El capítulo dedicado a la integración de la IA en procesos de investigación en instituciones educativas de Zacatecas, México, analiza, mediante un diseño empírico con enfoque cuantitativo, cómo docentes, investigadores y estudiantes se apropian de herramientas de IA generativa y qué riesgos y oportunidades perciben. A través de instrumentos aplicados a muestras de participantes distribuidos en distintos roles, se recaba información sobre las fases del proceso investigativo en las que se utiliza IA (búsqueda y revisión de literatura, formulación de problemas, diseño metodológico, análisis de datos, redacción y difusión), así como sobre niveles de aceptabilidad y preocupaciones éticas.

Los resultados muestran una fuerte inclinación a utilizar IA en tareas de apoyo textual, como la reformulación de párrafos, la generación de ejemplos o la organización preliminar de información bibliográfica. En cambio, se observa mayor cautela respecto de su uso para elaborar secciones completas de trabajos académicos, diseñar instrumentos de recolección de datos o realizar análisis estadísticos sin supervisión. Esta asimetría sugiere que los actores educativos reconocen la utilidad de la IA para aligerar cargas mecánicas, pero mantienen cierta reserva ante la delegación de tareas que consideran centrales para la formación de criterio investigador.

En el plano ético, el capítulo reporta que las principales preocupaciones mencionadas por los participantes se relacionan con la posible pérdida de pensamiento crítico y autonomía, la dependencia excesiva de recomendaciones automatizadas y la dilución de la autoría. Aspectos como el plagio, la originalidad o la confiabilidad de las respuestas de la IA también aparecen, pero no ocupan necesariamente el primer lugar, lo que evidencia una sensibilidad particular hacia la dimensión formativa de la investigación. El texto concluye que la incorporación de IA en la investigación educativa debe acompañarse de estrategias pedagógicas orientadas a fortalecer la capacidad de evaluar críticamente tanto las fuentes tradicionales como las generadas por algoritmos.

Desde la perspectiva de este capítulo de apertura, estas conclusiones permiten formular una advertencia más general: si la IA se integra de manera acrítica en procesos de formación e investigación, puede favorecer un desplazamiento de la atención desde la problematización y la construcción de sentido hacia la simple producción de textos y resultados *correctos*. Evitar este riesgo implica, precisamente, colocar la reflexión sobre los usos de IA en el centro de la agenda de desarrollo de capacidades científicas.

Inteligencia artificial generativa y ética profesional en la contaduría pública

El capítulo se sitúa en un campo donde la confianza social en el juicio experto es fundamental para la regulación económica y fiscal. A partir de una revisión documental de códigos de ética nacionales e internacionales, así como de literatura reciente sobre IA en contaduría y auditoría, los autores analizan las tensiones que emergen cuando herramientas capaces de generar textos, interpretar normas y sugerir estrategias se insertan en el quehacer profesional.

El análisis parte del reconocimiento de que la IA generativa puede apoyar en tareas como la elaboración de borradores de informes, la sistematización de cambios normativos, la simulación de escenarios fiscales y la preparación de materiales de comunicación con clientes. No obstante, el capítulo enfatiza que dichas funciones no son neutras, pues se sitúan en un marco de principios de ética profesional que incluye la integridad, la objetividad, la competencia profesional, la confidencialidad y la conducta profesional.

El texto identifica varios focos de tensión. Por un lado, la posible opacidad respecto del origen de los contenidos generados plantea interrogantes sobre la responsabilidad del contador en caso de errores o recomendaciones perjudiciales. Por otro, la facilidad con que la IA puede producir formulaciones verosímiles, pero no necesariamente rigurosas, puede conducir a un relajamiento en la revisión crítica de argumentos, sobre todo en contextos de alta carga de trabajo. Además, la introducción de IA como asistente en la toma de decisiones podría, si no se establece un marco claro, diluir la trazabilidad de las razones que sustentan un dictamen profesional.

Desde el punto de vista del presente capítulo, esta contribución subraya la importancia de que la ética profesional no se conciba como un conjunto de reglas estáticas, sino como un espacio de deliberación en el que deben incluirse explícitamente los problemas derivados de la mediación algorítmica. Ello implica revisar códigos, incorporar criterios sobre uso responsable de IA, establecer obligaciones de transparencia frente a clientes y autoridades, y fortalecer la formación en capacidades para auditar y cuestionar sistemas, más allá de la simple competencia en su operación.

Coaching empresarial y gestión administrativa

El capítulo de revisión sobre el uso de IA en el *coaching* empresarial compila literatura publicada principalmente en los últimos años, identificando tendencias en la utilización de

chatbots, plataformas de recomendación y sistemas de análisis de datos para apoyar procesos de desarrollo de liderazgo y acompañamiento organizacional. El texto muestra que la IA se presenta como una opción para escalar y abaratar servicios de *coaching*, al ofrecer interacciones automatizadas disponibles de forma continua y personalizadas a partir del análisis de cuestionarios, métricas de desempeño y otros datos.

Los autores advierten, sin embargo, que los estudios revisados señalan claras limitaciones: dificultad para manejar situaciones emocionalmente complejas, ausencia de empatía genuina, riesgo de respuestas estandarizadas poco sensibles al contexto y preocupaciones sobre privacidad y uso secundario de datos sensibles. Se observa una tendencia hacia modelos híbridos en los que la IA se utiliza como herramienta de apoyo (por ejemplo, para registrar y analizar patrones de interacción), mientras que la conducción de procesos transformacionales profundos permanece en manos de *coaches* humanos.

En paralelo, el capítulo sobre aplicación de IA en la gestión administrativa de micro, pequeñas y medianas empresas del sector construcción muestra cómo sistemas de IA pueden contribuir a optimizar procesos de administración, contabilidad, gestión de proyectos y toma de decisiones operativas en organizaciones con recursos limitados. A través de un estudio de caso o análisis empírico, se destaca que la adopción de estas tecnologías requiere no sólo inversión económica, sino también cambios en la cultura organizacional y en la capacitación del personal administrativo.

Ambos capítulos coinciden en una idea clave: la IA no reemplaza automáticamente la pericia humana, sino que redefine las fronteras entre lo que se automatiza y lo que se considera núcleo irreductible del juicio profesional. Este reordenamiento plantea preguntas que son, en última instancia, de carácter epistemológico y ético: ¿qué dimensiones del conocimiento experto son delegables?, ¿bajo qué condiciones?, ¿y quién asume la responsabilidad última por las decisiones compartidas entre agentes humanos y sistemas algorítmicos?

Comunicación científica, ciencia abierta y arquitectura editorial

Un cuarto eje relevante del libro se ocupa de los efectos de la IA en la comunicación científica y en las infraestructuras de ciencia abierta. Aquí la atención se desplaza desde el uso de IA en la producción de datos y análisis hacia su papel en la organización, evaluación y circulación del conocimiento.

Ciencia abierta, datos masivos e IA generativa

El capítulo *Hacia una nueva praxis de ciencia abierta dominada por datos masivos e inteligencia artificial generativa* examina cómo la convergencia entre grandes repositorios de datos, plataformas de acceso abierto e IA generativa está configurando una nueva ecología de la investigación. Los autores señalan que el ideal de ciencia abierta, promovido en las últimas décadas para democratizar el acceso al conocimiento, se ve atravesado por tecnologías que pueden facilitar esa democratización, pero también concentrar poder en actores que controlan infraestructuras y modelos.

Entre los beneficios, el texto destaca la capacidad de la IA para explorar vastos *corpus* de literatura, identificar patrones de citación, sintetizar hallazgos y generar representaciones accesibles de resultados para públicos no especializados. Asimismo, apunta que la IA puede ayudar a integrar datos heterogéneos de distintas fuentes, potenciando la investigación interdisciplinaria. Sin embargo, se advierte que la dependencia de plataformas específicas y de modelos de propiedad privada introduce asimetrías significativas entre instituciones que pueden costear servicios avanzados y aquellas que no.

El capítulo plantea que, en ausencia de políticas claras, la ciencia abierta corre el riesgo de degenerar en un modelo en el que los contenidos son nominalmente accesibles, pero su visibilidad y capacidad de ser encontrados y reutilizados depende de algoritmos opacos cuya lógica no es transparente para la comunidad científica. La necesidad de desarrollar marcos de gobernanza de datos y algoritmos en entornos de ciencia abierta se presenta, así, como un desafío ineludible.

Arquitectura editorial en salud y producción exponencial

El capítulo sobre el impacto de la IA generativa en la arquitectura editorial y el crecimiento exponencial de la producción científica en salud analiza los cambios que se están produciendo en los procesos editoriales. Se discuten usos de IA en tareas como la detección de similitud y posibles plagios, la sugerencia de revisores, la identificación automática de temas y palabras clave, y la asistencia a autores en la redacción y revisión lingüística de manuscritos.

El texto sostiene que estos procesos contribuyen a acelerar la circulación de artículos, pero también a incrementar el volumen global de publicaciones. En el ámbito de la salud, donde la presión por publicar resultados de investigaciones clínicas y epidemiológicas es

alta, la IA puede facilitar la aparición rápida de numerosos textos cuyo valor añadido es desigual. Esta situación exacerba tensiones conocidas, como la *inflación* de artículos y la dificultad de los equipos editoriales y de los lectores para discriminar contribuciones realmente significativas.

Desde una perspectiva crítica, el capítulo plantea que la IA no sólo automatiza partes del circuito editorial, sino que también puede influir en qué temas, enfoques y tipos de artículos reciben mayor visibilidad, en función de cómo se configuren los algoritmos de recomendación y búsqueda. Se sugiere que, para preservar la integridad y la diversidad del ecosistema científico, es necesario discutir abiertamente el uso de IA en revistas y plataformas, establecer políticas de transparencia sobre su intervención en procesos editoriales y redefinir criterios de calidad que no se apoyen exclusivamente en métricas cuantitativas.

Ética, marcos regulatorios y justicia algorítmica

La dimensión ética y normativa de la IA ocupa un lugar destacado en el libro, especialmente en los capítulos dedicados a la justicia algorítmica, a la revisión de marcos regulatorios y éticos sobre IA en la investigación científica y al análisis del uso práctico y la transición regulatoria de la IA en el derecho.

Justicia algorítmica y gobernanza ética

El capítulo *Justicia algorítmica y gobernanza ética ante los sesgos de la inteligencia artificial* ofrece una revisión crítica de literatura académica e institucional sobre sesgo algorítmico, discriminación automatizada y marcos de gobernanza. A partir de estudios ya clásicos sobre racismo algorítmico y de análisis de documentos de organismos internacionales como la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, por sus siglas en inglés), la Organización para la Cooperación y el Desarrollo Económicos (OCDE) o la Comisión Europea, las autoras argumentan que los sistemas de IA no son neutrales, sino que inscriben en su funcionamiento estructuras de poder y prejuicios históricos presentes en los datos con los que se entrenan.

El capítulo subraya que los sesgos algorítmicos tienen características que los vuelven especialmente problemáticos: se reproducen a gran escala, se presentan con una apariencia de objetividad técnica y son difíciles de detectar para los sujetos afectados. Ejemplos de sistemas de puntaje de riesgo de reincidencia, de reconocimiento facial o de filtrado de

candidatos en procesos de contratación muestran cómo decisiones aparentemente *eficientes* pueden traducirse en violaciones sistemáticas de derechos.

Frente a ello, las autoras abogan por una concepción robusta de justicia algorítmica que no se limite a la corrección técnica de modelos, sino que incorpore principios de transparencia, rendición de cuentas, participación de grupos afectados y protección de derechos humanos. Señalan que, mientras en algunos contextos se ha avanzado en marcos regulatorios, en países como México persiste un vacío normativo específico sobre IA y discriminación tecnológica, lo que incrementa la vulnerabilidad de poblaciones ya de por sí marginadas.

Desde la perspectiva de este capítulo de apertura, esta contribución es clave porque traslada discusiones globales al contexto latinoamericano y porque evidencia que la investigación científica que utiliza IA no puede desentenderse de la dimensión de justicia. Modelos entrenados con datos biomédicos, educativos, laborales o ambientales pueden reproducir sesgos de origen si no se desarrollan mecanismos de revisión y participación multidisciplinaria.

Marcos regulatorios y ética de la inteligencia artificial en la investigación científica

Complementariamente, el capítulo *Entre principios y práctica: revisión de marcos regulatorios y éticos sobre inteligencia artificial en la investigación científica* realiza un análisis de documentos normativos y de líneas guía emitidos por organismos internacionales, agencias nacionales y asociaciones profesionales. A partir de un enfoque documental cualitativo, los autores identifican principios recurrentes (beneficencia, no maleficencia, autonomía, justicia, explicabilidad, privacidad, responsabilidad) y examinan en qué medida se traducen en obligaciones concretas para investigadores e instituciones.

El capítulo destaca la brecha entre formulaciones de alto nivel de carácter muchas veces aspiracional y la realidad de prácticas cotidianas en las que no siempre existen procedimientos claros para evaluar riesgos de proyectos que incorporan IA. Se observa, por ejemplo, que comités de ética en investigación biomédica o social no siempre cuentan con criterios específicos para valorar el uso de algoritmos en recolección, análisis y almacenamiento de datos, ni para exigir transparencia sobre modelos utilizados.

Este análisis conduce a la propuesta de fortalecer capacidades regulatorias y de revisión ética incorporando expertos en IA en comités, desarrollando guías específicas

para proyectos algorítmicos y promoviendo la formación de investigadores en aspectos de protección de datos, equidad algorítmica y explicabilidad. De esta forma, la ética de la investigación se actualizaría para responder a retos que no estaban presentes cuando se formularon muchos de los marcos vigentes.

Inteligencia artficial en el derecho y transición regulatoria

El capítulo profundiza en un campo donde las decisiones tienen consecuencias directas sobre derechos fundamentales y donde el uso de herramientas de IA se expande rápidamente. El texto describe aplicaciones como el análisis masivo de jurisprudencia, la predicción de probabilidad de éxito de demandas, la generación de borradores de contratos y escritos y el uso de sistemas de apoyo a la decisión judicial.

El análisis muestra que, aunque estas herramientas prometen eficiencia y consistencia, también plantean desafíos profundos: riesgo de pérdida de independencia judicial si se otorga demasiado peso a recomendaciones algorítmicas, opacidad en la forma en que los modelos ponderan distintos factores y posibilidad de consolidar prácticas discriminatorias si los datos de entrenamiento reflejan decisiones pasadas sesgadas. El capítulo revisa iniciativas regulatorias en distintas jurisdicciones y concluye que aún no existe un consenso claro sobre los límites aceptables del uso de IA en procesos judiciales.

Para la investigación científica, este escenario es relevante por dos razones. Primero, porque estudios empíricos sobre IA y derecho se convertirán, previsiblemente, en insumo para la elaboración de futuras normativas, lo que coloca a la comunidad científica en una posición de co-responsabilidad en la configuración del marco regulatorio. Segundo, porque los propios proyectos de investigación que utilicen datos jurídicos o trabajen con IA en este campo deberán someterse a estándares de protección de derechos especialmente exigentes.

Epistemología, cultura y co-creación humano-máquina

Finalmente, varios capítulos adoptan un enfoque filosófico, estético o cultural para discutir el estatuto epistémico de la IA generativa y las implicaciones de la co-creación entre humanos y algoritmos. Entre ellos se encuentran los textos sobre el estatuto epistémico de la IA generativa, la escucha humana e IA en investigación musical, la evolución de la representación gráfica y las experiencias de co-creación crítica en educación superior.

Conocimiento o apariencia

El capítulo *Conocimiento o apariencia: el estatuto epistémico de la inteligencia artificial generativa* formula una pregunta fundamental: ¿pueden los *outputs* de sistemas generativos considerarse conocimiento, o son sólo apariencias de conocimiento? Para responder, el autor recurre a la tradición epistemológica que define conocimiento como creencia verdadera y justificada, y examina cómo se transforman estos términos cuando el productor de contenidos es un sistema estadístico que carece de creencias en sentido estricto.

El texto argumenta que los sistemas de IA generativa producen, ante todo, configuraciones lingüísticas que tienen una alta probabilidad de ser coherentes y verosímiles, dadas las distribuciones de palabras en sus corpus de entrenamiento. Estas salidas pueden ser útiles para explorar espacios discursivos, formular hipótesis o sintetizar información, pero no garantizan, por sí mismas, ni verdad ni justificación. La justificación sigue dependiendo de prácticas humanas de verificación empírica, contraste argumentativo y revisión por pares.

Desde esta perspectiva, el capítulo propone distinguir entre productividad textual y avance epistémico. La IA puede multiplicar la primera sin incrementar necesariamente el segundo. De ahí que, en el contexto de la investigación científica, sea crucial no confundir la facilidad para generar textos con la producción de conocimiento nuevo. Esta distinción ilumina varios de los casos discutidos en otros capítulos, en los que la IA se utiliza tanto para explorar datos como para redactar manuscritos.

Escucha, representación y co-creación en educación superior

Los capítulos sobre escucha humana e IA en investigación musical y sobre la evolución de la representación gráfica desde el trazo humano al algoritmo abordan la IA en tanto mediadora de prácticas sensibles que involucran percepción, cuerpo y creatividad. En el ámbito musical, se analiza cómo herramientas de IA pueden asistir en el análisis de patrones sonoros, en la generación de variaciones y en la exploración de estructuras musicales, pero también cómo estas mediaciones plantean límites cuando se trata de capturar dimensiones de la escucha vinculadas a la experiencia subjetiva y la intersubjetividad.

En el caso de la representación gráfica, se estudia la transición histórica desde técnicas manuales hasta sistemas generativos capaces de producir imágenes a partir de instrucciones textuales. El capítulo examina cómo esta transición reconfigura la relación entre idea, gesto

y resultado visual, y discute la tensión entre la desmaterialización del proceso creativo y la persistencia de valores asociados al trazo humano.

Por su parte, el capítulo *Entre el prompt y el diseño: experiencias de co-creación crítica humano-inteligencia artificial en educación superior* describe experiencias pedagógicas en las que estudiantes y docentes utilizan sistemas generativos como herramientas de co-diseño, reflexionando explícitamente sobre las decisiones creativas que toman y sobre la influencia del algoritmo en la estética de los productos finales. La noción de *co-creación crítica* que se propone enfatiza que el valor pedagógico de estas experiencias radica menos en el producto y más en el proceso reflexivo de negociación entre intenciones humanas y sugerencias algorítmicas.

En conjunto, estos capítulos muestran que la IA no sólo transforma la forma en que se investiga y se comunica, sino también la manera en que se imagina lo que significa crear, interpretar y conocer. Al llevar los límites de la automatización al terreno de la sensibilidad y la estética, obligan a repensar categorías como autoría, originalidad y experiencia estética en clave postdigital.

Síntesis crítica: convergencias, tensiones y horizontes

Leído como un todo, el libro configura un mosaico donde la IA aparece simultáneamente como tecnología de frontera, problema social, oportunidad pedagógica, amenaza ética y objeto de reflexión filosófica. De este mosaico pueden extraerse algunas convergencias y tensiones que ayudan a orientar futuras investigaciones.

Convergencias

Desplazamiento del foco desde *herramienta* a *ecosistema socio-técnico*: varios capítulos coinciden en que hablar de IA implica hablar de regulaciones, infraestructuras, culturas organizacionales, dinámicas de poder y marcos éticos que la hacen posible o la limitan.

Centralidad de la formación y el pensamiento crítico: tanto en investigación educativa, contaduría, *coaching* o salud, se resalta la necesidad de formar profesionales capaces de interrogar algoritmos, no solo de operarlos, y de preservar la autonomía de juicio frente a recomendaciones automáticas.

Potencial epistémico en dominios complejos: casos como paleogenómica, estratificación de riesgo clínico o detección de plagas muestran que la IA puede abrir preguntas y patrones que exceden las capacidades humanas en términos de escala y velocidad.

Tensiones

Eficiencia *vs*. justicia: la misma tecnología que optimiza procesos puede reproducir desigualdades si se alimenta de datos sesgados o se implementa sin mecanismos de control y reparación.

Productividad textual vs. valor epistémico: la IA generativa facilita la producción masiva de textos académicos, pero esto no se traduce necesariamente en mayor calidad de conocimiento; al contrario, puede diluir estándares y saturar el espacio editorial.

Democratización vs. concentración: mientras algunos capítulos celebran la escalabilidad y accesibilidad de servicios mediadas por IA, otros alertan sobre la concentración de poder en pocas plataformas y la dependencia de infraestructuras propietarias.

Horizontes para la investigación científica con inteligencia artificial

A partir de los aportes del libro, se pueden delinear al menos tres líneas estratégicas para reorientar la relación entre IA e investigación científica:

1. *Diseñar marcos de gobernanza integrados,* que articulen políticas de adquisición, adaptación y asimilación de tecnologías de IA con estándares éticos y regulatorios específicos para la investigación, incluyendo criterios de transparencia en el uso de IA en todas las etapas del ciclo científico.

2. *Reconfigurar la formación de investigadores*, incorporando competencias en alfabetización algorítmica, auditoría de modelos, gestión crítica de datos y reflexión epistemológica sobre el estatuto del conocimiento mediado por IA.

3. *Impulsar investigaciones interdisciplinarias*, que traten a la IA no sólo como herramienta instrumental, sino como objeto de estudio que requiere la colaboración de ciencias sociales, humanidades, ciencias naturales, ingeniería y artes para entender su impacto integral.

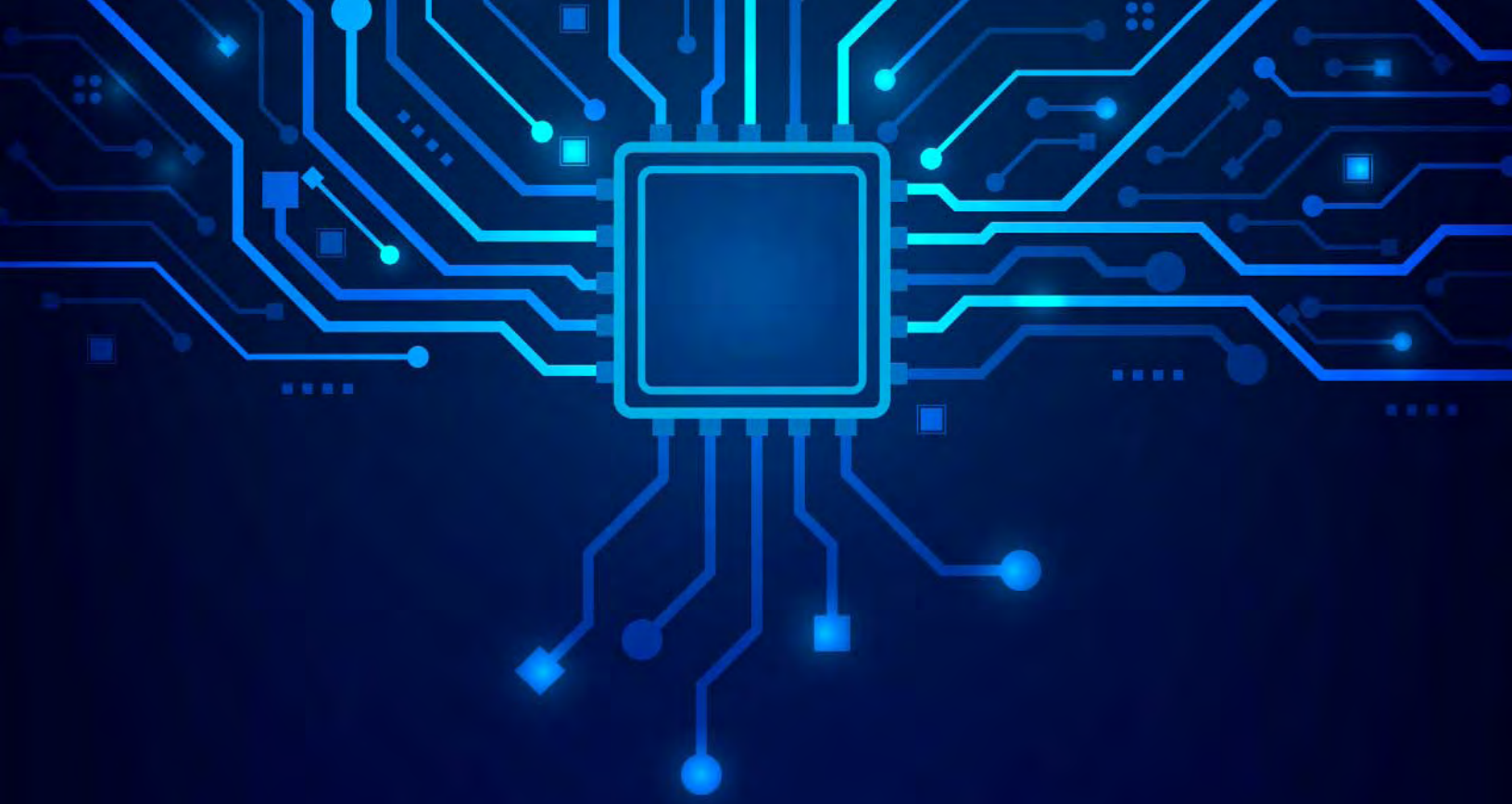
Conclusiones

El libro *Inteligencia artificial: experiencias y reflexiones sobre la investigación científica* no es sólo una compilación de casos de uso de IA en distintos campos, sino una invitación a repensar de manera radical la relación entre tecnología, conocimiento y sociedad. Los capítulos que lo componen muestran, desde ángulos diversos, que la IA está reconfigurando las prácticas científicas de manera profunda, y que esta reconfiguración no es un proceso técnico aislado, sino un fenómeno socio-técnico atravesado por relaciones de poder, valores culturales, marcos normativos y concepciones de justicia.

El análisis transversal propuesto en este capítulo ha permitido identificar cinco grandes tensiones que emergen de las experiencias y reflexiones documentadas: la tensión entre eficiencia y justicia, entre productividad textual y avance epistémico, entre democratización y concentración de poder, entre automatización y autonomía profesional, y entre innovación metodológica y reproducción de sesgos estructurales.

Frente a estas tensiones, el libro no ofrece respuestas definitivas, pero sí establece algunos puntos de partida imprescindibles para la deliberación colectiva. En primer lugar, subraya que la adopción de IA en la investigación científica debe ser un proceso reflexivo, contextualizado y sujeto a evaluación continua, no una carrera acrítica por la *innovación*. En segundo lugar, destaca la importancia de fortalecer capacidades de gobernanza ética y regulatoria, tanto a nivel institucional como a nivel de las propias comunidades científicas. En tercer lugar, insiste en que la formación de investigadores debe incorporar, de manera transversal, la capacidad de interrogar sistemas algorítmicos, de reconocer sus límites y de preservar la autonomía de juicio crítico.

Finalmente, el libro plantea que cualquier proyecto de investigación con IA implica, en última instancia, una toma de posición sobre qué entendemos por inteligencia, conocimiento, creatividad y justicia. Estas categorías no pueden quedar definidas de facto por las capacidades de los sistemas disponibles, sino que deben ser objeto de deliberación pública, interdisciplinaria y democrática. En este sentido, *Inteligencia artificial: experiencias y reflexiones sobre la investigación científica* constituye una contribución valiosa no sólo para comprender el presente de la IA en la ciencia, sino para imaginar y construir futuros más justos, equitativos y epistémicamente responsables.



01.

Condiciones sociales en la planeación estratégica para la adquisición de inteligencia artificial en las organizaciones

José Antonio Cisneros Jiménez
Universidad Nacional Autónoma de México
ORCID: 0009-0006-1880-5204

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.01

01.

Condiciones sociales en la planeación estratégica para la adquisición de inteligencia artificial en las organizaciones

Introducción

Desde los albores de la administración, como disciplina, se ha entendido la tecnología como el eslabón que, ubicada en el trasfondo de los medios y los fines para la gestión de las organizaciones, enlaza a los seres humanos con las técnicas hacia el mundo social y el mundo organizacional que gestionan. La comprensión de la tecnología ha evolucionado con el tiempo. Hoy en día, el disruptivo desarrollo de frontera de los algoritmos, del *software*, de dispositivos y de *hardware*, hace que las tecnologías de la información de vanguardia se transfiguren y sean etiquetadas, genéricamente, como inteligencia artificial (IA).

En general, la connotación del concepto *inteligencia artificial* refleja nociones muy dispersas y alejadas de una posible definición formal. La IA puede ser, tanto una técnica, como un producto; una herramienta o un proceso, según se quiera asociar la idea a un tipo de comportamiento inteligente, autónomo o simplemente a la noción de un desarrollo innovador como valor agregado de un producto.

Si bien la evolución del desarrollo tecnológico avanza de la mano con la del ámbito económico, no se puede decir lo mismo de la comprensión social de éste que, más allá de la noción de que las comunidades de conocimiento tienen la ambición de precisar con certeza la disrupción de la frontera tecnológica y su impacto social, inexorablemente se diluye esta meta ante la dificultad que posa el adquirir, comparar, dialogar y participar activamente en la formación del conocimiento de algo nuevo tanto para un individuo como para la sociedad en su conjunto. La difusión superficial de los medios de comunicación, la incesante mercadotecnia, la rápida evolución tecnológica, el despliegue de los productos y servicios de las grandes empresas y gobiernos, no han hecho sino añadir vaguedad y ambigüedad al concepto IA, lo que afecta, indudablemente, cómo reaccionan los individuos y las organizaciones frente a la oportunidad o el riesgo que implica la IA en sus actividades críticas.

Discernir la diferencia entre una oportunidad y un riesgo es propio de una exploración estratégica del entorno para la planificación de una administración inteligentemente informada; no reactiva, sino activa. Asimismo, esclarecer el plan de acción para la adquisición de una tecnología de IA, su adaptación al entorno operativo requerido, y su asimilación por parte de los trabajadores, requiere un trabajo cuidadoso por parte de los gestores de tecnología; implica una estrategia para transferir la tecnología efectivamente a la organización.

El problema de la transferencia tecnológica (TT) se desenvuelve en dos extremos: el de la comprensión de las fuerzas activas del entorno de las organizaciones; y el de la planeación estratégica. Ambos frentes están relacionados intrínsecamente por aquello que cuestionan; el valor y el sentido de las preguntas de uno son las respuestas del otro. Así, la problematización de las fuerzas del entorno alrededor de los desarrollos de IA se analiza desde la perspectiva de los planteamientos discursivos que describen las condiciones de posibilidad políticas-regulatorias, técnicas-tecnológicas y sociales-culturales, para enfrentarlas con los requerimientos de planificación estratégica para la adquisición, adaptación y asimilación de soluciones de IA en las organizaciones.

El planteamiento problemático de los discursos sobre la inteligencia artificial

Si uno trata de construir una definición de la IA a partir de sus términos componentes, como se hizo con la aproximación del concepto de *planeación estratégica*, se llegará a una proposición algo distante de lo que en un primer momento se pueda atribuir concepción cultural. Esto es debido a que, independientemente de la postura teórica a la que se quiera llegar a enfrentar, la polisemia de la IA está dominada en todos los campos epistémicos tanto por la ambigüedad técnica como por la vaguedad cultural asociada al término.

Los problemas de ambigüedad se producen porque las palabras no se usan en todos los contextos o situaciones humanas para connotar las mismas propiedades que se refieren, y se podría decir que *la referencia de una expresión es ambigua*. En el caso de la IA, ocurre cuando se habla de *inteligencia*, sin definir bien del todo las condiciones para considerarse como inteligencia dada la noción de un comportamiento inteligente o a un conjunto de líneas de código específicas que designan a un algoritmo específico como tal.

Por otro lado, en los casos de vaguedad la cuestión no es que un término tenga varias connotaciones, sino que, dada una connotación, hay casos que nos plantean dudas

porque no tenemos claro dónde termina el campo de aplicación de la palabra; se podría expresar que *el sentido de una expresión es vago*. Un ejemplo en la IA se presenta cuando el objeto al que se le confiere la facultad de inteligente, como un sistema de reconocimiento facial, un automóvil que se conduce automáticamente, un brazo robótico, o la interfaz de un dispositivo (Ludeña, 2015).

Por ejemplo, la Organización para la Cooperación y el Desarrollo Económicos ([OECD], 2019) describe a un sistema de IA compuesto de tres elementos principales: 1) sensores para recopilar datos sin procesar del entorno; 2) lógica operativa para interpretar los datos para un conjunto determinado de objetivos y proporcionar un resultado de recomendaciones, predicciones o decisiones; y 3) actuadores para cambiar el entorno en función de ese resultado. Podemos contrastar algunas de definiciones más y distinguir aquellas que versan entre un enfoque prescriptivo y uno descriptivo de la IA. En el enfoque prescriptivo (normativo), la relación propuesta es la de fin/medio (teleológica), un *deber ser* orientado por objetivos; en la orientación descriptiva (positiva), la correspondencia está dada entre causa/efecto, en la cual *las cosas son como son* (Betancur et al., 2022). Algunos de éstos se presentan en la Tabla 1.

Tabla 1

Ejemplos de definiciones de la IA según el ámbito de despliegue social

Ámbito	Postura descriptiva	Postura prescriptiva
Tecnología	"La inteligencia artificial es el campo científico de la informática que se centra en la creación de programas y mecanismos que pueden mostrar comportamientos considerados inteligentes. En otras palabras, la inteligencia artificial es el concepto según el cual las máquinas piensan como seres humanos… un sistema con inteligencia artificial tiene la capacidad de llevar a cabo procesos que realizaría una persona inteligente." (Porcelli, 2020, p.56)	"AI is the intelligence of machines or software and is also a branch of computer science that studies and develops intelligent machines and software. This field of computer science, AI, is defined as "the study and design of intelligent agents" where an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success. It is also defined as "the science and engineering of making intelligent machines" (Patterson, 1992, p.2)

Tabla 1

Ejemplos de definiciones de la IA según el ámbito de despliegue social

Ámbito	Postura descriptiva	Postura prescriptiva
Trabajo	"Puede ser una inteligencia predictiva, prescriptiva o descriptiva en el uso de plataformas como la Inteligencia artificial, que resulta útil para labores propias de recursos humanos, supervisión de rendimiento o micro gerencia. También puede contener una inteligencia afectiva para el análisis de los recursos humanos, o incluso una inteligencia de asistencia o colaborativa" (Granados, 2022, p.119)	"AI refers to machines that imitate some aspects of "human intelligence, such as perception, learning, reasoning, problem-solving, language interaction and creative work […] Currently, no AI system can be considered generally 'intelligent' in the sense that it can perform well in different contexts (so-called strong AI), which is an ability of human intelligence." (Stolpe & Hallström, 2024, p. 3)

Las aristas de la IA son muchas y deben atenerse a diversos rasgos sociales, económicos, regulatorios, filosóficos, tecnológicos, psicológicos, entre otros muchos más (Alegría, 2022). En el esfuerzo por delimitar la comprensión del concepto y el entorno de los desarrollos tecnológicos de la IA, se tiene que hablar sobre cómo se pueden plantear estrategias organizacionales problematizándola críticamente contra sus narrativas y discursos. ¿Cómo formular una estrategia de TT para la adquisición de una solución de IA para una organización? El entorno de las organizaciones abarca más que el mero mercado y el desenvolvimiento de una dinámica de competencia económica por generar utilidades. Los conflictos de los diversos frentes sociales a los que una organización interpela con sus planes estratégicos demandan dar respuesta a una problemática persistente: el dinamismo de los cambios sociales en toda su amplitud.

Conviene mejor pensar a la sociedad en términos de *campos problemáticos*, en tanto sugieren el planteamiento de un problema y el trabajo continuo para ofrecer una respuesta, en lugar de representar a la sociedad como un sector fragmentado institucionalmente. La idea sería plantear el posicionamiento estratégico haciendo énfasis, no en el antagonismo competitivo entre actores, sino en las fuerzas sociológicamente problematizadas; es decir, como un conjunto de condiciones de posibilidad que facilitan u obstaculizan el posicionamiento estratégico de la organización en un contexto social específico.

Retomando la clasificación de las escuelas de la planeación de Mintzberg et al. (1998), la premisa central de la escuela del posicionamiento abarca la formulación de estrategias que se basa en la teoría de que para una empresa determinada hay sólo unas pocas posi-

ciones a su favor en el mercado (Sergio, 2011, p. 59). Minztberg sugiere una clasificación de tres principales perspectivas, que abarcan a diez escuelas de pensamiento: la Escuela de Diseño, Escuela de Planificación, Escuela de Posicionamiento; Escuela Emprendedora, Escuela Cognitiva, Escuela de Aprendizaje, Escuela de Poder, Escuela Cultural, Escuela Ambiental, Escuela de Configuración. Para la escuela del posicionamiento, lo que interesa es la ventaja competitiva, que se puede definir como el medio por el cual una empresa puede aumentar y mantener las ganancias a lo largo del tiempo, y esta ventaja puede ser de dos tipos: por costos o por diferenciación (Grant, 2021).

Porter (2008) ofrece un *framework* para evaluar la estructura del mercado dada una industria específica, dadas cinco categorías o fuerzas que determinan el estado de la competitividad en éste: la amenaza de nuevos participantes, el poder de negociación de proveedores y clientes, la amenaza de sustitutos y la rivalidad competitiva dentro de la industria. Un *framework* puede entenderse como un conjunto de conceptos delimitado por categorías que hacen una estructura de elementos destinada a la implementación práctica, para la construcción o diseño de soluciones de problemas, es decir, su enfoque es uno analítico o técnico. Sea pues un conjunto de preguntas dirigidas para la formulación de un plan estratégico, éste se considera un *framework*

Este *framework* permitiría construir una narrativa de la estructura industrial y del mercado de la IA, buscando responder y dar solución al problema de la competitividad organizacional. La TT para adquirir la IA, requiere de otras respuestas porque hace otras preguntas; su problematización es distinta: cuestiona si las capacidades de la organización son adecuadas para adquirir, adaptarse y asimilar tecnología que se considera como una oportunidad para la organización. Por tanto, se requiere de la evaluación crítica de las posturas, narrativas y factores sociales específicos que se dan en el entorno de las organizaciones, más allá de las dinámicas del mercado. Para el caso que compete a esta investigación, el núcleo de los discursos de la IA responde al planteamiento de un campo problemático alrededor de las condiciones de posibilidad sociales de la IA. Las preguntas que surjan del planteamiento de esos problemas podrán servir como el *framework* para problematizar la estrategia de TT.

Por problematización se puede entender un proceso del pensamiento crítico que evalúa las condiciones de posibilidad, es decir, la validez y los límites para cuestionar un fenómeno de interés. El planteamiento de problemas implica más que el de formular preguntas y afirmar discursos, pues, como señala Sánchez y López (2011):

> "la cuestión aquí radica en comprender que la simple interrogación no puede suplantar el verdadero pensamiento acerca del [concepto en cuestión] debido a que ya se está presumiendo que esa realidad exterior existe en principio con lo cual sólo queda la respuesta afirmativa y, en su caso, un agregado para dar cuenta de sus principales rasgos distintivos, o sea, su adjetivación" (p. 131).

Un problema debe ir más allá de la oposición entre la corrección de lo verdadero y lo falso, lo que exige ir más allá de un mero reconocimiento por la existencia o presencia del objeto o fenómeno al que se dirige el cuestionamiento en el afuera de la conciencia. Es decir, si efectivamente existe o no como un concepto reflejado en lo real del mundo. En el caso de IA, se encuentran con facilidad discursos interesantes y *verdaderos*, pero no un planteamiento de problema. Es decir, una forma más elevada del pensamiento que supere la dicotomía de lo útil y lo inútil de una formulación estratégica por la competitividad, que eluden la posibilidad de darle un auténtico sentido estratégico, tal que se logre: "adelantar signos que nos lancen hacia un aprendizaje, ya que es entonces cuando puede acceder a enunciar verdades, o mejor aún que sus actos de verdad se expresen en los problemas mismos, puesto que es necesario construirlos" (Sánchez & López, p. 131).

La formulación de estrategias requiere de un esfuerzo creativo formidable y la fuerza del pensamiento es indiscutiblemente importante para lograrlo. Desde siempre, a la administración, en su postura más pragmática, se le ha considerado como arte, después de todo. Y el arte de formular estrategias exige de la creatividad del estratega para efectuarse y consolidarse efectivamente en una genuina problematización para cumplir con su propósito organizacional. El entorno de las organizaciones se considera contingente y azaroso; nunca existirá una solución administrativa inmutable y verdadera frente al dinamismo de lo social. Por eso, el estratega requiere ser flexible en su crítica para establecer la perspectiva con la cual ir al encuentro con *el afuera*. En efecto, ¿Es posible encontrar una forma de relación positiva del pensamiento con lo desconocido o con lo no pensado, que dé cuenta del acto del pensamiento estratégico? Se advierte que ya no es suficiente con enunciar la exterioridad de la relación con un discurso, sino que ha llegado el momento de crear una nueva posición estratégica; del terreno conceptual de su planteamiento central emanará el sentido de las nuevas preguntas del nuevo *framework* estratégico pensado.

Discusión crítica y construcción de nuevos enfoques estratégicos

Siguiendo la lógica discutida, se proponen tres perspectivas que atienden problematizaciones específicas de los campos sociológicos de interés para la estrategia de adquisición tecnológica de IA. Con campos sociológicos se hace referencia al recorte epistémico de categorías funcionales que se puede efectuar conceptualmente sobre la categoría mayor y abarcadora de *lo social*. Esto hace eco de las ideas de Niklas Luhmann, por ejemplo, quien describió a la sociedad como un sistema compuesto por múltiples subsistemas, como pueden ser el político, el económico, el jurídico, el religioso, el artístico, o el científico.

Las tres perspectivas que se proponen son: *política-regulatoria del desarrollo de TIC*; *técnica-tecnológica del progreso tecnológico*; y *social-cultural del desarrollo humanista*. Tras la lectura analítica de cuatro tesis doctorales que abordan cada temática desde una postura crítica con gran detalle, se presenta la clasificación de las subtemáticas de interés para la derivación lógica de la centralidad de los cuestionamientos respectivos a cada uno sobre las condiciones de posibilidad del cambio social desde la IA. El criterio para la clasificación de éstos se da a partir de la identificación en los textos de la presencia de las tensiones de un discurso dominante y su evaluación crítica de una alternativa de reactiva o sumisa; así como la crítica del cambio social desde una postura activa o creadora. Cada una de las subtemáticas que se presentan en la Tabla 2 se describen en términos de dos conceptos que representan posturas enfrentadas o complementarias, desarrolladas a detalle en los textos de los que se ha extraído la respectiva crítica temática.

Tabla 2

Elementos dialécticos de las posturas críticas frente a la IA para la problematización de las condiciones de posibilidad del cambio social

Perspectiva analítica	Política-regulatoria del desarrollo de TIC	Técnica-tecnológica del progreso tecnológico	Social-cultural del desarrollo humanista
Discursos dominantes	Globalización de las TIC: agenda neo-republicana & sociedad	Determinismo tecnológico en la IA: instrumentalización vs. agencias	El pensamiento humanista en la IA: escatologías vs. cosmopolitismo
Crítica reactiva	Disrupción de la IA: Desregulación vs. institucionalización polisémica	Posthumanismo y transhumanismo: mente y cuerpo	Sociedad de la información y el conocimiento

Tabla 2

Elementos dialécticos de las posturas críticas frente a la IA para la problematización de las condiciones de posibilidad del cambio social

Perspectiva analítica	Política-regulatoria del desarrollo de TIC	Técnica-tecnológica del progreso tecnológico	Social-cultural del desarrollo humanista
Crítica Activa	Condiciones históricas del nacimiento del internet: comunitarismo & liberalismo	Condiciones técnicas para la existencia de la IA: funcionalismo vs. inteligencia	Condiciones Sociales de la aceptación social de la IA: sumisión vs. resistencia

La perspectiva crítica *política-regulatoria del desarrollo de las TIC* es estudiada por Cruz (2017) y abarca el cuestionamiento de las temáticas del nacimiento del internet, donde se complementan las posturas del comunitarismo y del liberalismo tecnológico (pp. 49-87); y del de la globalización de las TIC, que discute el alcance de la agenda neo-republicana y la aceptación social de las TIC (pp. 91-143), indudablemente el soporte de las plataformas de la IA; el aporte la tesis de Alegría (2022) completa la tercer temática, la disrupción regulatoria de la IA, oponiendo y ejemplificando la desregulación del desarrollo tecnológico y la contingente institucionalización polisémica que se da frente a la dificultad de precisar el alcance de la IA como tecnología para las aplicaciones políticas.

La postura *técnica-tecnológica del progreso tecnológico* se encuentra en las páginas de tres trabajos de investigación; primero, el libro de Parente (2010), donde presenta las posturas filosóficas y técnicas para interpretar a los artefactos tecnológicos y la técnica en sí misma hasta las implicaciones de cada una de ellas como para designar el cambio tecnológico como determinista o abierto a una complejidad mayor, siendo las principales posturas enfrentadas la del carácter instrumentalista o protésico del artefacto contra la de las teorías de la agencia material; la tesis de Barrios (2019, pp. 219-303), explora la temática de los fundamentos metafísicos y tecnológicos en común encontrados en las corrientes contemporáneas de posthumanismo y el transhumanismo, analizando sus implicaciones tanto para el control y el desarrollo de dispositivos análogos al funcionamiento de la mente y cuerpo; finalmente, la tesis de López (2013, pp. 344-417) explora las condiciones técnicas y científicas para la existencia de la IA desde la postura de un funcionalismo psicologista y reduccionista, enfocado sólo en el poder de la mente, contra aquellos teóricos que analizan la imposibilidad de replicar completamente la inteligencia por medio de las técnicas actuales, tanto en el caso de la IA simbólica como sub-simbólica.

Finalmente, el enfoque *social-cultural del desarrollo humanista*. La tesis Barrios (2019) estudia la convergencia del pensamiento humanista en la IA con una tradición metafísica de la técnica de la esencia y la perfección que se imbuyen en el pensamiento tecnológico, casando los conceptos del progreso tecnológico de corte cosmopolita a las escatologías históricas como el *mileniarismo* (pp. 29-123); el mismo autor la desarrolla en su tesis de la sociedad de la información y el conocimiento, desde una postura genealógica de la teoría de sistemas para explicar cómo deviene la información en la sociedad a ser el epígono del poder para todos los sujetos que le poseen (pp. 123-208); por último, la tesis de López (2013, pp. 417-448) también desarrolla una crítica a la temática de las condiciones sociales de la aceptación social de la IA, oponiendo una alternativa de sumisión social opacada por la cultura de masas contra una posible resistencia activa contra las narrativas dominantes de los grupos de poder, como empresas globales que dominan la vanguardia en los desarrollos de punta de la IA.

Se propone analizar el enfoque estratégico de la TT desde la postura de la adquisición de un bien o servicio tecnológico. Para esto, se diferencian tres etapas de consecución respecto a sus metas y objetivos como parte de una estrategia de adquisición tecnológica. En ese sentido, es de interés definir los conceptos de la adquisición, adaptación y asimilación. La adquisición tecnológica es el proceso en el cual una organización empresa buscará adquirir tecnologías que estén disponibles en ese momento en el mercado, que sean de calidad probada, que garanticen la solución a sus problemas o necesidades, que cuesten lo menos posible y que impliquen el menor número de restricciones o limitaciones (Velásquez & Medellín, 2005, p. 13).

Por su parte, se habla de adaptación tecnológica cuando se requiere cambiar el diseño y características de una tecnología que ha sido desarrollada para operar en condiciones muy distintas a las de la región donde se va a utilizar la tecnología que se adquiere (Velásquez & Medellín, 2005, p. 35). La asimilación tecnológica se refiere a una profundización del conocimiento en el uso de la tecnología, incrementando notablemente su avance en la curva de aprendizaje; es el proceso en el que las funciones técnicas dirigidas al objetivo de producir un bien o un servicio dentro de la organización se realicen con mayor eficiencia, debido a que cuentan con la mejor y más actualizada información y que disponen de conocimientos adecuados y oportunos. Consta de tres aspectos principales: documentación, capacitación del personal, actualización (Velásquez & Medellín, 2005, p. 36).

Situados desde cada uno de los planteamientos problemáticos abordados en esas temáticas y subtemáticas de las condiciones de posibilidad del cambio social de la IA, se derivará un posiblemente cuestionamiento respectivo para la formulación de TT para la adquisición de soluciones de IA. Esto, retomando la clasificación de los enfoques estratégicos específicos de la administración, y también considerando su convergencia con los procesos de adquisición en la estrategia de adquisición tecnológica. El *framework* resultante de la congruencia entre los cuestionamientos de interés para la TT, el enfoque estratégico organizacional y la evaluación del cambio social de la IA se da en la Tabla 3. Este nuevo *framework* es el resultado del trabajo intelectual hecho sobre el análisis del planteamiento de las problemáticas de los trabajos citados en cada caso, presentando un concepto central de interés estratégico y matizado por la caracterización del proceso de TT. La estrategia general de adquisición de tecnología se presenta en el flujo inferior del esquema, que puede aplicar tanto para la tecnología de IA como para la evaluación estratégica de otras tecnologías.

Tabla 3
Marco analítico para la problematización del entorno social de la IA y su TT

Proceso de la TT	Entorno Social	Planeación Estratégica
Estrategia de adquisición: ¿La complejidad del paquete tecnológico se acopla a las capacidades y recursos organizacionales para decidir si adquirirla por: 1) compra completa o llave en mano, 2) desarrollar por copia un elemento del sistema, o 3) emprender una solución innovadora?	Regulación tecnológica: ¿Cuáles son las regulaciones vigentes que hay en el país para transferir la tecnología?	Ventaja competitiva: ¿El uso de la tecnología ha justificado las mejoras en las empresas del sector que la utilizan?
	Vinculación Institucional: ¿Qué instituciones facilitan la adquisición de conocimiento sobre el uso y el impacto de la tecnología en la región?	Madurez Tecnológica: ¿La innovación tecnológica se encuentra en una etapa incipiente y cuánto tiempo se proyecta tarde en madurar?
	Ética Organizacional: ¿Existen recomendaciones internacionales que orienten el alcance del uso, control y desarrollo de la tecnología en las organizaciones?	Fuerzas del mercado: ¿Cuáles son los canales y los agentes que comercializan la tecnología?

Tabla 3
Marco analítico para la problematización del entorno social de la IA y su TT

Proceso de la TT	Entorno Social	Planeación Estratégica
Estrategia de Adaptación: ¿La información de especificaciones y requerimientos de los elementos paquete tecnológico son suficientes para adaptar: 1) su instalación a la infraestructura de la organización, 2) operarse eficientemente, 3) expandir los procesos de operación a otras rutinas y problemas?	Funcionalismo técnico: ¿La tecnología pretende aplicarse como una herramienta complementaria o suplementaria al trabajador?	Eficiencia operativa: ¿La tecnología se puede acoplar a los procedimientos y procesos de la organización para realizar más efectivamente este trabajo?
	Ergonomía y riesgos: ¿Cuáles son los riesgos de averías que pueden suscitarse para que el trabajador evite hacer un uso técnico incorrecto o indebido de la tecnología?	Flexibilidad creativa: ¿La tecnología permite adaptarse a la aplicación de nuevos problemas organizacionales?
	Dependencia técnica: ¿Cuáles son los riesgos psicológicos y sociales frente al abuso o corrupción en el uso de la tecnología?	Soporte y confiabilidad: ¿Es posible reconfigurar, reparar, sustituir los elementos y funciones del sistema tecnológico con otros compatibles en el caso de fallas o de alta demanda?
Estrategia de Asimilación: ¿Las capacidades organizacionales al operar el paquete tecnológico son suficientes para asimilar: 1) el aprendizaje de su aplicación; 2) la generación de nuevo conocimiento; 3) la gestión del nuevo conocimiento?	Accesibilidad y transparencia: ¿el manejo las interfaces es claro y sencillo de usar para los usuarios y desarrolladores?	Dinámicas de poder: ¿Qué escenarios de resistencia existen frente a la aceptación o rechazo del uso de la nueva tecnología?
	Valores Culturales: ¿Cómo se perciben y generan las opiniones de las personas de dentro y fuera de la organización hacia el uso de esta tecnología?	Liderazgo y coordinación: ¿Cuáles son las estrategias para implementar y sostener un cambio favorable a los planes organizacionales?
	Difusión Pública: ¿Las condiciones socio-técnicas de la organización permiten establecer una buena comunicación abierta y vinculada con los agentes del entorno externo?	Gestión del conocimiento: ¿Los flujos de datos, información y conocimiento son gestionados para toda la organización o sólo para los canales precisos, tal que haya fugas de conocimiento desconocidas?

Conclusiones

La IA requiere ser considerada más que una mera tecnología útil en la búsqueda de soluciones organizacionales; donde hay valor hay poder y, ante éste no se puede accionar estratégicamente ni resistir coordinadamente, si se desconoce el entramado de intereses sociales que hay alrededor de los discursos de la tecnología. La aceptación social de una tecnología, y en especial con la difusión masiva que las TIC han acarreado globalmente

con creces desde su gestación, implica muchas veces considerar a la sociedad como un agente paroxísticamente pasivo y sumiso, un consumidor frente a las decisiones y el poder de decisión con el que las instituciones reguladoras y empresas desarrolladoras de estas tecnologías dirigen el rumbo del desarrollo social, con la afianzada potestad de *poner las reglas del juego tras el telón de la competencia económica*.

No es del todo así y, para comprender cómo es que la valoración y el lugar del poder con el que se identifica e interpreta la acción de los agentes dominantes y el rol de todos los demás, hay que retomar posturas que permitan dar sentido al diseño estratégico de un plan de TT para la adquisición de la IA o de cualquier tecnología; debe cuestionarse críticamente al entramado social desde la política, técnica y la filosofía. Y esto sólo puede realizarse desde un planteamiento adecuado de los problemas que ahonde en el fondo de cada cuestión: las condiciones de posibilidad del cambio social.

El resultado de ese esfuerzo del pensamiento estratégico demuestra que en el correlato de la disciplina administrativa hay sub-campos teóricos y enfoques estratégicos más adecuados para, a su vez, plantear cuestionamientos precisos a cada etapa de un proceso de adquisición tecnológica, su adaptación y su asimilación a las organizaciones. Queda por mencionar que todo ese esfuerzo es efectuado por la pericia de un estratega que sabe reconocer y adaptarse con flexibilidad creativa al problema concreto de pensar el futuro dadas las condiciones de posibilidad de su presente. La presencia del estratega, del gestor de la tecnología en todo caso, es quien da seguimiento a la ejecución y control de los planes; después de todo, la planeación es sólo una parte del arte del saber cómo administrar.

Referencias

Alegría, J. (2022). *Desafíos filosóficos e implicaciones morales de las narrativas sobre la inteligencia artificial* [Tesis de Doctorado de la Universidad de Salamanca, España]. Repositorio institucional. https://gredos.usal.es/bitstream/handle/10366/150978/PDF_PiedraAlegriaJJ_IA.pdf?sequence=28

Barrios, F. (2019). *Transhumanos, superhombres y máquinas humanas: Nietzsche contra la metafísica* [Tesis de Doctorado de la Universidad Complutense de Madrid, España]. Repositorio institucional. https://hdl.handle.net/20.500.14352/17135

Betancur, H., Soto, N., & Vargas-González, C. (2022). Análisis sociológico del sistema contable desde Niklas Luhmann. *Revista Mexicana de Sociología*, *84*(2), 417-446. https://doi.org/10.22201/iis.01882503p.2022.2.60283

Cruz, M. (2017). *Las percepciones políticas y tecnológicas de los desarrolladores de tecnologías de la información y la comunicación móvil como indicadores del cambio social: el caso de España* [Tesis de Doctorado de la Universidad Complutense de Madrid, España]. Repositorio institucional. https://hdl.handle.net/20.500.14352/15476

Granados, J. (2022). Análisis de la inteligencia artificial en las relaciones laborales. *Revista CES Derecho*, *13*(1), 111-132. https://revistas.ces.edu.co/index.php/derecho/article/view/6395

Grant, R. (2021). *Contemporary strategy analysis*. John Wiley & Sons.

López, M. (2013). *Inteligencia artificial: condiciones de posibilidad técnicas y sociales para la creación de máquinas pensantes* [Tesis de Doctorado de la Universidad Complutense de Madrid, España]. Repositorio institucional. https://hdl.handle.net/20.500.14352/38189

Ludeña, L. (2015). Vaguedad, ambigüedad y cadenas de comunicación. *Revista de Teoría del Derecho de la Universidad de Palermo*, *2362*(3667), 65-82. https://www.palermo.edu/derecho/pdf/teoria-del-derecho/n3/TeoriaDerecho_03.pdf

Mintzberg, H., Ahlstrand, B., & Lampel, J. (1998). *Strategy safari: A guided tour through the wilds of strategic mangament*. Free Press.

OECD (2019). *Artificial Intelligence in Society*. OECD Publishing. https://doi.org/10.1787/eed-fee77-en

Parente, D. (2010). *Del órgano al artefacto: acerca de la dimensión biocultural de la técnica*. Memoria Académica.

Patterson, D. (1992). *Introduction to Artificial Intelligence and Expert Systems*. Prentice-Hall of India.

Porcelli, A. M. (2020). La Inteligencia Artificial y la Robótica: sus dilemas sociales, éticos y jurídicos. *Derecho Global. Estudios sobre Derecho y Justicia*, *6*(16), 49–105. https://doi.org/10.32870/dgedj.v6i16.286

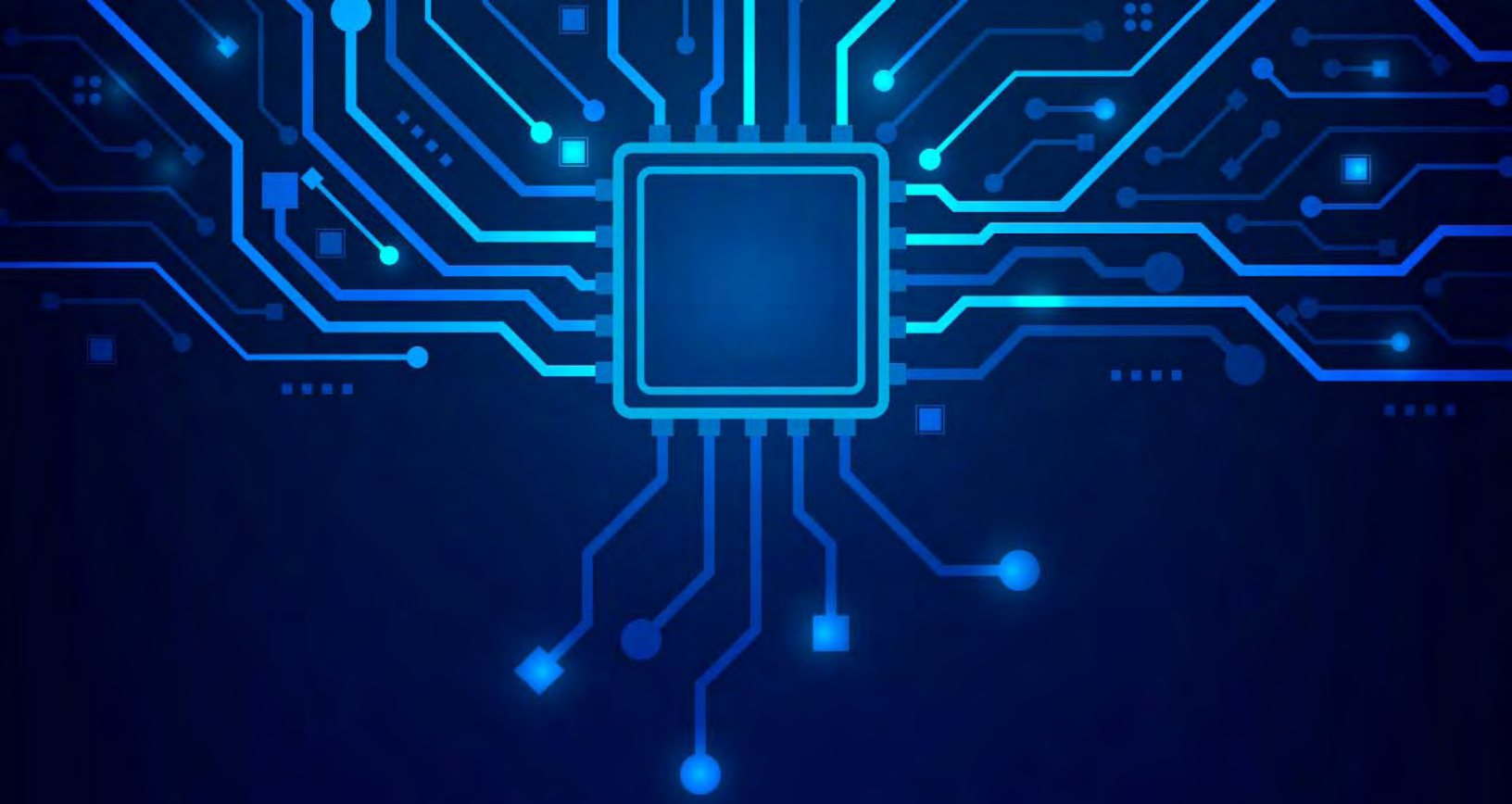
Porter, M. (2008). *The five competitive forces that shape strategy.* Harvard Business Rreview, 86(1), 78.

Sánchez, J., & López, S. (2011). Problematización de la responsabilidad social o colectiva. *CIENCIA ergo-sum, Revista Científica Multidisciplinaria de Prospectiva*, *18*(2), 126-134. https://www.redalyc.org/pdf/104/10418753003.pdf

Sergio, J. (2011). *Strategic Management: The theory and practice of strategy in (business) organizations*. Kgs. Lyngby. DTU Management, 1.

Stolpe, K., & Hallström, J. (2024). Artificial intelligence literacy for technology education. *Computers Education Open*, 6, 100159. https://doi.org/10.1016/j.caeo.2024.100159

Velásquez, G., & Medellín, E. (2005). *Manual de transferencia y adquisición de tecnologías sostenibles.* CEGESTI, U.S. Agency for International Development Guatemala – Central American Program. https://www.sica.int/busqueda/busqueda_archivo.aspx?Archivo=-libr_9388_1_16062006.pdf



02.

Employment Situation for Recent University Graduates in Mexico City (2020-2024)

Mariajosé López Laiza
Instituto Politécnico Nacional
ORCID: 0009-0001-7126-1430

Rita Ávila Romero
Instituto Politécnico Nacional
ORCID: 0000-0002-4214-8105

Cesaire Chiatchoua
Instituto Politécnico Nacional
ORCID: 0000-0001-8915-7562

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.02

02.

Employment Situation for Recent University Graduates in Mexico City (2020-2024)

Introduction

Higher education is responsible for creating and offering young people, students, specialized technical knowledge and skills, which allow their insertion into the professional sectors of the industry in a simpler way. We must consider the changing needs of the labor market, influenced by innovation and competitiveness, at a global level; requiring study centers that their programs and curricula evolve and adapt along with the market, seeking to ensure that graduates have the knowledge, skills, and competencies required by the demand of employers.

This shows that higher education plays a fundamental role in the development of society, both at the personal, social, economic and individual levels. It should be mentioned that, over the years, higher education has become a key tool for the training of future professionals, and that this is highly trained, allowing them to face the challenges of the world of work, which is constantly changing and in the process of globalization. Often, as economies transform and technological advances transform the nature of work, the relationship between the labor market and higher education becomes closer and more complex.

In this regard, Miranda Quiñones et al. (2016), argue that the changes derived from digitalization have transformed the labor market, higher education will have to intensify the generation of professionals, adjusting to the new requirements, to achieve adequate labor insertion.

Consequently, the labor market is influenced by higher education; the academic training of applicants directly affects the demand and supply of skills and knowledge, which influences the employment rate, salary levels and quality of life of professionals. The link between these two worlds is due to advances in educational training, as well as the dynamics

of job requirements, and these can feed back through a dynamic of constant, interrelation and adjustment between sectors.

This shows a relationship between the labor market and the education sector. This link has been studied by several authors. López Paniagua and Becerra-Medina (2018) underline the contradiction between educational policy and its programs that encourage the link between supply and demand for professionals. This contradiction is expressed, on the one hand, in the promotion of policies to modernize the education sector, and on the other, the growing contraction of the labor market for university graduates.

Regarding the above, Camarena Gómez and Velarde Hernández (2010) mention that the processes of linkage and social relevance between universities and the labor market have not been very promising in recent decades, mainly in those countries that, due to their structural deficiencies, continue to focus their educational policy on only increasing enrollment and terminal efficiency. One of the possible solutions to the contradictions is to formulate proposals for the generation of strategies that link the public, private, and university sectors to reduce institutional gaps, according to González Rodríguez et al. (2022).

Following the above, Salas Durazo and Murillo García (2013) conclude that education systems, higher education institutions, and government policies must work together in favor of young people to guarantee access to relevant and quality training, which allows them to respond to and face the demands of the labor market. which is dynamic and competitive.

In Mexico, mainly in Mexico City, the higher education system presents challenges such as school dropout, failure, low terminal efficiency, low academic performance, career changes, among others, which in turn does not combine with the demands of the labor market. Therefore, the objective of this research is to demonstrate the relevance of academic training at the time of job placement of recent university graduates in Mexico City. We start from the fact that an outdated curriculum will lead to shortcomings in the academic training of graduates and will complicate their entry into the labor market.

After the introduction, section one, literature review, analyzes some currents of thought on the labor market, the situation of young people in the Mexican labor market, and higher education in the country. Section two presents the methodology of this research guided by secondary data. Section three shows the results of the study. Finally, a proposal of strategies and the conclusions of the study.

Literature review

In this section, several topics are addressed. It begins with some currents of thought on the labor market, then the situation of young people in the Mexican labor market and finally, higher education in the country.

Currents of thought on the labour market

A theory that analyzes the dynamics of overeducation and the knowledge gap is not granted within the literature on the labor market. Although there are certain theoretical developments that can be used as a frame of reference. The most relevant is the Human Capital Theory, proposed by Gary Becker; defines human capital as the set of skills and qualities that individuals have to increase their productivity; the main element mentioned is knowledge, but other factors ranging from a sense of punctuality to someone's state of health are also considered, they are also relevant according to the *Harvard Business Review* (2017).

Paul Romer, his most relevant work was developed in the field of economic growth and his research shows that the accumulation of ideas sustains long-term economic growth (De La Cruz & Gallardo, 2024), once again brought growth into view in the field of theoretical research that gives rise to endogenous growth models (Jiménez, 2010), the theory of endogenous growth explains why the economies of industrialized countries generate much higher amounts of per capita than they did 100 years earlier.

Next, the heterogeneous theory of knowledge mentions that the dynamic that exists between the level of schooling and the delay in knowledge is not so serious. People with the same level of schooling have a diversity of abilities; The scenario they plan to find is workers with a higher level of formal education but with skills that are at the same level as workers with a lower level of schooling; and this group of workers would be called "overeducated", this due to the years of formal education training, but they would be considered with the appropriate education to occupy the job due to their skills and abilities, denied that they do not have marketable skills according to Burgos Flores & López Montes (2011).

Subsequently, the assignment theory states that jobs with different requirements are assigned heterogeneous workers. Work is not homogeneous, together with the assumption that the theory of human capital handles, which also contemplates investment in human capital and the requirements or characteristics that work demands, this justifies the different salaries that workers obtain, even though they have the same academic or educational level.

This does not guarantee that they will be the same, as it will depend on the functions they perform at work. It is when the existence of workers' overeducation is raised, described as the scenario in which the level of education is higher than that required by the work that the individual will occupy, and this is due to the fact that the skills and knowledge acquired by the worker in his training do not coincide with those required for the job according to Burgos Flores (2008).

Finally, the Knowledge Economy, created by Kenneth Arrow and Fritz Machlup, is defined as the economic sector that makes use of information and knowledge in order to generate value, and thus be able to offer people services and products that improve their lives. In a concrete way, it could be defined as encouraging development based on the promotion of knowledge and the reduction of the increase in traditional resources, such as machinery, infrastructure, among others; defined as an investment in intangible capital, which has not stopped its growth since the end of the twentieth century according to Navarro Martínez (2018).

Young people in the Mexican labor market

According to López et al. (2021), recent graduates of universities or higher levels in Mexico highlight the problem that has brought them, the lack of broad and diverse knowledge that allows them to have a good development in the labor field, they relate it to the lack of experience. In Mexico there is a considerable number of graduates and the great demand to occupy the jobs that are offered in the labor market of professionals, bringing with it the consequence of the decrease in salaries. This has had the effect that graduates have had to adapt to working in various areas of companies, however, these do not correspond to their professional profile, or it is not the most suitable for them.

For Viveros and Cortés (2023), the implementation of new technologies in industries rethinks the skills necessary to join the labor market, since advanced cognitive skills and adaptation to the labor market they wish to enter will also be demanded. This new implementation in industry eliminates jobs in some sectors but compensates for them with the net increase in jobs in other sectors.

Higher Education in Mexico

Mexico had not contemplated this new interaction of industry and new technologies for its future development, now we must place ourselves in the year 2021, because after

40 years the new Higher Education Law is approved, specifically the General Law of Higher Education (LGES, by its acronym in Spanish).

With this change, the national scenario opens to meet current needs at the international level in the technological, economic, social and educational fields through a higher education system that opens transformation due to the new demands of industry. Higher Education is an asset with a strategic purpose, which together with the new law allows it to live up to international practices and standards in the fields of academic, professional, scientific, research and development of technology according to Viveros and Cortés (2023).

In Mexico, having a professional career is not a guarantee of a job immediately after graduation. “For several years it has been identified that the annual growth of graduates from higher education is greater than the annual growth of the national economy and that around 80% of the total number of professionals is concentrated in 20 careers” (Viveros & Cortés, 2023); showing that our labor market is incapable of offering opportunities.

For a transformation in Mexico a participatory democracy is necessary, and that this is a reality and a necessary basis. For this, it is the pedagogical relationship between the people and the government, which begins in school and the training that is given in them. That is why the New Mexican School is proposed, which will work and function due to the 2022 curriculum, which is described as a creative, complex and unfinished idea or work. A greater relationship between the school system, life and work is necessary; because what is sought is that it does not remain as an academic exercise. The most relevant point for this work is to establish a conscious educational system so that the educational actions and experiences of democratic movements can be resumed in accordance with the Morenista National Convention (2024).

Method or research

This study analyzes the elements that are generated between the formation of an academic profile and the work environment faced by graduates in public universities in Mexico, particularly in Mexico City. For González Rodríguez et al. (2022), this research is presented as a contextualization of the Mexican environment from the perspective of employment and the labor competencies demanded by the labor market, together with the framework of public action, which is assumed as an integral element of the labor market-university-society triad.

This research is descriptive and analytical, they are presented as secondary data derived from the Instituto Nacional de Estadística y Geografía (INEGI) and DATA MEXICO sources with information from the 2020 Population and Housing Census, the panoramas of the labor market and graduates, in order to establish the bases of a subsequent applied study with a focus on the training trajectories of the students, and to generate strategies that are integrated into the plans, programs and curricula of the universities, as part of their institutional policy agenda. Various reports prepared by national and Mexico City bodies were analyzed with the aim of collecting the necessary elements for the description of the employment panorama and graduates in Mexico City.

Results: current situation of the labor market in Mexico City

Since the global COVID-19 pandemic, there has been an increase in the negative impact on employment and economic growth at the national level, which has caused uncertainty about the future of employment that has been affected by the digital revolution, the automation of processes and the disintegration of value chains in a large part of the sectors of the economy. A negative effect of the crisis has been the increase in unemployment, which has highlighted the need to invest in employment, as well as in the professional skills and abilities of young people that will be necessary in the future according to Blázquez et al. (2020).

As mentioned above, there are factors that the digital revolution has brought with it; that have affected certain sectors of the economy have caused the disappearance of jobs, the transformation of existing jobs offered by companies, as well as the needs and knowledge required to fill these new jobs. These new changes require new professional competencies, and the gap between the necessary and existing competencies has grown considerably in recent years according to López et al. (2021).

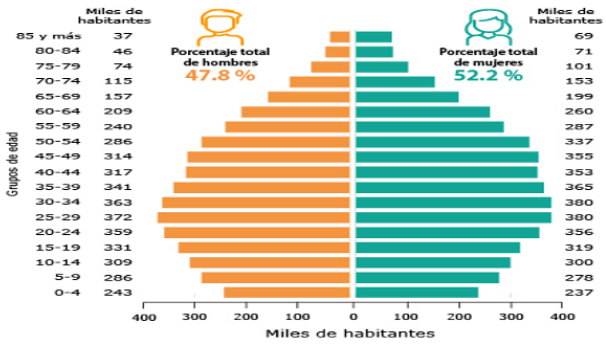
Graduates encounter their problems, which ranges from unemployment, underemployment, very low wages and no benefits; this, together with competitiveness and the lack of jobs in the labor market, makes it difficult to obtain jobs. They will encounter complications when entering the labor field and performing their profession, since companies seek to have their employees with experience and avoid job training expenses for employers (López et al., 2021).

Now it is at the moment where companies set their conditions for the hiring of the services of recent graduates, they are looking for those individuals who can perform qualified

jobs, and the reality is that there is a lack of individuals with these characteristics and this is because the educational system in Mexico has not been able to act and give a response at the height of the changes presented, causing a gap of professional competitiveness, so if measures are not taken soon, it will simply grow more in the coming years according to Blázquez et al. (2020). Figure 1 shows the Number of Inhabitants. Mexico City in 2020 had 9,209,944 inhabitants, of which 4,805,017 are women and 4,404,927 are men. Mexico City ranks 2nd nationally in number of inhabitants by state.

Figure 1

Number of inhabitants by age and sex

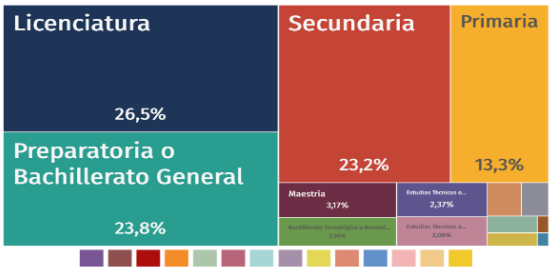


Note. National Institute of Statistics and Geography (INEGI) with information from the 2020 Population and Housing Census.

The level of schooling is defined as the highest level of education approved by the population aged 3 years and over according to the National Education System or its equivalent in the case of studies abroad (Figure 2).

Figure 2

Levels of schooling of the population aged 15 and over in Mexico City (Distribution of the total population)



Note. DATA MEXICO with information from the 2020 Population and Housing Census.

The Economically Active Population (EAP) refers to the population aged 15 and over that had a link to economic activity or that sought it in the reference week. The labor participation rate refers to the percentage of the population aged 15 and over that is within the EAP (Figure 3).

Figure 3

Evolution of the EAP in Mexico City

Note. DATA MEXICO with information from the National Survey of Occupation and Employment (ENOE).

For the second quarter of 2024, the labor participation rate in Mexico City was 62.9%, showing an increase of 1.45 percentage points compared to the previous quarter (Figure 4).

Figure 4

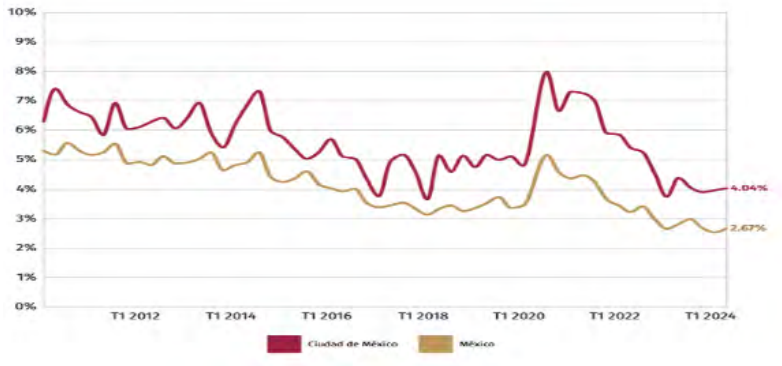
Evolution of the employed population in Mexico City (differentiating between formal and informal workers)

Note. DATA MEXICO with information from the National Survey of Occupation and Employment (ENOE).

The Employed Population refers to the Economically Active Population (EAP) that carries out an economic activity. The employment rate refers to the percentage of the EAP that is employed. For the second quarter of 2024, the employed population in Mexico City was 4.83 million people, being 2.53% higher than the previous quarter (Figure 5).

Figure 5

Evolution of the unemployment rate in Mexico City

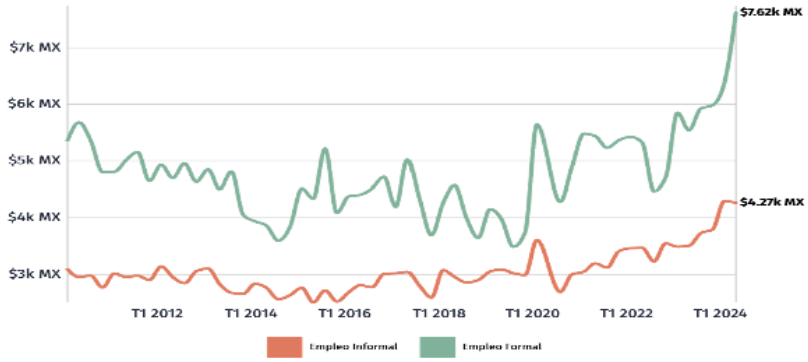


Note. DATA MEXICO with information from the National Survey of Occupation and Employment (ENOE).

For the second quarter of 2024, the unemployment rate was 4.04% (204 thousand people), showing an increase of 0.069 percentage points compared to the previous quarter. The average salary is the average amount of money that the worker receives in a given country or region over a period of time (Figure 6).

Figure 6

Evolution of average monthly wages in Mexico City (differentiating between formal and informal workers)



Note. DATA MEXICO with information from the National Survey of Occupation and Employment (ENOE).

For the second quarter of 2024, the average monthly salary of $6.1 thousand MX is $691 MX higher than the previous quarter.

Mexico City has a population of 9,209,944 inhabitants. The population belonging to the 20 to 24 age group corresponds to 359 thousand men and 356 thousand women: this being one of the largest segments of the population in the city. It should be mentioned that with respect to the level of schooling of the population aged 15 and over, in Mexico City only 26.5% have completed a bachelor's degree, this being the highest percentage of studies in the population. The population employed in formal jobs for the first quarter of 2024 is approximately 2,643,578 inhabitants and the unemployment rate observed is 4%, almost double that observed nationally in Mexico.

Mexico City has a high level of incurrence of studies at the bachelor's level, which is to say that its population mostly reaches the degree or studies the university, reflecting that it has a high professional population, and this can be seen reflected in the formal employment of the city and its population. Although it is worrying, the fact that, and the unemployment rate is so high relative to the national level.

Proposals and strategies

For Attwell and Hughes (2019), when talking about education and employment, decisions about these should be placed in the context of the labor market, since many times young people's decisions are related to the structure of particular opportunities, guiding their decisions and aspirations due to the understanding they have of the structures.

Santamaría et al. (2022) point out that labor markets are increasingly unstable, which has led the system to the development and creation of new labor market information systems, segmenting individuals into multiple groups or segments. This brings with it problems and challenges in general, in relation to the allocation of jobs and the use of data in education and employment environments. The labor market is currently highly competitive and is constantly evolving, which is why companies when looking for human capital demand that it be constantly updated and be prepared to face challenges at the level of skills, transversal and technical.

The strategy that should be chosen is more than clear: the trends are leaning towards training, which is the right alternative to respond to the demands and needs of the labour

market. In the first instance, it is necessary to investigate and analyze what the demands and needs of the labor market are, and not simply to train individuals at the level of professional qualification; before creating training and education programs and plans. Communication is essential so that the relationship that exists between the educational training of the supply is coherent with the demand of the business field, and this must occur in the educational and labor fields, since they share common interests, and this makes it easier for them to work in the same direction.

Among the proposals mentioned, greater investment in the field of education could be highlighted to improve insertion in the labor market, through incentives such as scholarships, so that they can continue with their studies, such as master's degrees or postgraduate degrees, the increase in places due to careers with lower employability and, finally, the promotion of aid or bonuses to employers to encourage the hiring of women in jobs and encourage equality (Jiménez-García, 2020). In the case of Mexico, to obtain improvements, we must mention that there is a relationship between the performance of the labor market, social deprivation, and the evolution of prices.

The unemployed population is an indicator that can show the economic growth scenario that the country is going through, and being a relevant variable, an increase in it contributes to increasing the number of individuals belonging to the poverty segment. Mexico seeks to reduce the population in poverty through programs that encourage formal employment, since this will allow individuals to have: a home, an income, security and social protection; with this allowing higher rates of economic growth to be achieved, accompanied by internal market dynamics, price stability, and sustained and healthy growth. Informal employment and its reduction are considered a necessary measure to counteract poverty, this could be worked on through tax incentives and credit support, encouraging and inviting fiscal formalization in a gradual and successful manner, through a transition that is regulated by the public policy framework or a program according to Attwell and Hughes (2019).

According to the Foro Institucional (2024), the labor market is undergoing constant transformation; projections indicate that by 2027, approximately 40% of workers' core skills will be disrupted. Technology has outpaced the capacity of firms to design and implement training programs for their employees. While Artificial Intelligence is a primary driver of this disruption, other factors such as the green transition and global geopolitical and economic conditions are also directly impacting job roles—driving the rise of emerging occupations while putting others at risk of displacement.

The lack of economic opportunity is becoming one of the most significant risks for the coming years. In the long term, barriers such as restricted economic mobility could intensify, leaving large segments of the population without access to opportunities. Consequently, upskilling and the development of advanced competencies will be essential tools to guarantee an equitable labor market in the future.

During the Forum's Annual Meeting in Davos, under the theme 'Creating Jobs and Growth for a New Era' (Foro Institucional, 2024) several key points were addressed: generative AI is expected to increase productivity unevenly in 2024. While it is projected to boost innovation and productivity in high-income countries, its impact in low-income nations remains a subject of consideration. Furthermore, digital employment continues to expand; by 2030, the number of digital jobs globally is expected to reach approximately 90 million. These roles tend to be higher-paying, help mitigate skilled labor shortages in high-income economies, and provide greater opportunities for young workers in low-income countries. If managed effectively, this could leverage global talent pools and foster new pathways for economic growth across all nations, regardless of their income level.

Conclusions

When a restructuring of the higher education system is put on the table, it is that currently having a professional career is no longer enough, nor is it a guarantee to occupy a place in the labor market after graduation. In Mexico City, work could be done on the education side with the government, institutions and the General Law of Higher Education; on the industry side with the companies that are in the area; together with a projection and strategies of new trends, such as: artificial intelligence, digital jobs, the reduction of unemployment, the increase in the importance and relevance of skills to be able to obtain a certain job.

Taking advantage of the advantages of location, diversification, industry and training that Mexico City has as the capital of Mexico, being a perfect place for there to be a good development of the relationship between education and the labor market; with this, allowing other states to replicate the operation in the future and this brings improvements at the national level and we are a competent country at the international level.

Without leaving aside that the improvements, modifications and adaptations of the curriculum of young people should always be constantly updated because the industry demands in terms of technologies and skills required to occupy jobs by young university

graduates. Who aspire to integrate into the labor market, with the aim of personal growth; that entails economic growth, well-being, stability, which in a certain way is given by a good salary or salary or remuneration for their work, which is the reflection of their training, skills and abilities as an individual in the place of their work. By having a new generation, which has better job opportunities, it will allow, in the future, the macroeconomic aspect to improve, reducing unemployment in their employment, improving the living conditions of families, greater Gross Domestic Product, among other details.

References

Attwell, G., & Hughes, D. (2019). Learning about Careers: Open data and Labour Market Intelligence. *ITEN Ibero-American Journal of Distance Education*, *22*(1), 81. https://doi.org/10.5944/ried.22.1.22289

Blázquez, M. L., Masclans, R., & Canals, J. (2020). *The professional skills of the future: a diagnosis and an action plan to promote youth employment after COVID-19*. IESE Business School University of Navarra. https://www.iese.edu/es/wp-content/uploads/sites/2/2020/07/IESE-Competencias-profesionales-del-futuro-2020.pdf

Burgos Flores, B., & López Montes, K. (2011). Effects of overeducation and knowledge gap on salaries and job search of professionals. Results of a study based on opinions and perceptions of university graduates and employers. *Educational Profiles*, *33*(134). https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982011000400003#:~:text=En%20tanto%2C%20la%20teor%C3%ADa%20del,un%20abanico%20diverso%20de%20capacidades

Burgos Flores, B. (2008, October). On education and knowledge gap in the labor market of professionals. *Journal of Higher Education*, 148. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-27602008000400004

Camarena Gómez, B. O, & Velarde Hernández, D. (2010). Higher education and the labour market: social linkage and relevance Why? and What for? *Social Studies Journal of Contemporary Food and Regional Development*, *17*(1Esp.), 105-127.

Morenista National Convention. (2024, February 12). *10. Education for the Revolution of Consciences*. https://morenademocracia.mx/proyecto-de-nacion-2024-2030/10-educacion-para-la-revolucion-de-las-conciencias/

De la Cruz, G., & Gallardo, A. L. (Eds.). (2024). *Education in the 2018-2024 six-year term*. Views from educational research. https://doi.org/10.22201/iisue.9786073095501e.2024

Foro Institucional. (2024, 10 septiembre). *6 tendencias laborales para observar en 2024*. Foro Económico Mundial. https://es.weforum.org/stories/2024/02/6-tendencias-laborales-para-seguir-en-2024/

González Rodríguez, G. I., Tovar Vergara, M., & Vargas Hernández, J. G. (2022). The Crossroads Between the Labor Market, Universities, and Student Graduation: An Exploratory Study. *Revista Andina de Educación*, *6*(1), e201. https://doi.org/10.32719/26312816.2022.6.1.2

Harvard Business Review. (2017, August 6). Gary Becker and human capital. *La Nación*. https://www.lanacion.com.py/harvard-business-review/2017/08/06/gary-becker-y-el-capital-humano/

Jiménez, F. (2010). *Elements of macroeconomic theory and policy for an open economy*. Pontificia Universidad Católica del Perú. http://files.pucp.edu.pe/departamento/economia/LDE-2012-02a.pdf

Jiménez-García, J. R. (2020). And after university what? Analysis of the transitions from the education system to employment in Spain. *International Journal of Sociology*, *78*(3), 162. https://doi.org/10.3989/ris.2020.78.3.18.133.

Llamas, R. V., & Hernández, J. M. O. (2020). Multidimensional poverty and the labor market in Mexico. *Cuadernos de Economía*, *39*(79), 139-165. https://doi.org/10.15446/cuad.econ.v39n79.71201

López, V. M., Sánchez Nájera, S. A., & González Olmos, C. (2021). Difficulty in employing young graduates with no experience. *Humanities, Technology and Science of the National Polytechnic Institute*, 25, 1. https://revistaelectronica-ipn.org/ResourcesFiles/Contenido/26/HUMANIDADES_26_001009.p

Miranda Quiñones, A. E., Gómez Moreno, J. J., Gómez Miranda, V. D., & Lizama, L. (2016). The Labor Market and its Relationship with Higher Education. *Journal of Science and Engineering of the Higher Technological Institute of Coatzacoalcos*, *3*(3), 12-17.

Navarro Martínez, J. Á. (2018, October 17). *The knowledge economy*. McGraw Hill–Economics. https://www.mheducation.es/blog/la-economia-del-conocimiento

López Paniagua, R., & Becerra-Medina, M. S. (2018). Higher education and the labor market in Mexico: contextualization of their linkage. In Amador, R., Ceballos, R. & Villa Lever, L. (Coords.). *The challenges of education. Vol. XIV of The Social Sciences and the National Agenda. Reflections and proposals from the Social Sciences*. Consejo Mexicano de Ciencias Sociales, A.C.

Santamaría, O. Q., García-Álvarez, J., & Del Carmen Santos-González, M. (2022). Analysis of the effectiveness of vocational training plans for employment from the point of view of the labour market. *Revista Fuentes*, *1*(24), 90-103. https://doi.org/10.12795/revistafuentes.2022.15258

Salas Durazo, I. A., & Murillo García, F. (2013). University professionals and the Mexican labor market: convergences and asymmetries. *Journal of Higher Education*, *XLII*(1), (165), 63 – 81. http://www.redalyc.org/articulo.oa?id=60428314004

Viveros, C. A. H., & Cortés, M. E. (2023). Labor market for professionals in Mexico in 2022: analysis using a hierarchical multi-criteria model. *Figures Revista Académica de Investigación*, *4*(3), 119-140. https://doi.org/10.22201/fesa.26832917e.2023.4.3.276

03.

Inteligencia artificial y paleogenómica predictiva: Incompatibilidad Rh y Kell en el colapso demográfico neandertal

Luis Ramón Carreño Durán
Instituto Politécnico Nacional, México
ORCID: 0000-0003-1815-0907

Aura Patricia Hernández Olicón
Instituto Politécnico Nacional, México
ORCID: 0000-0002-4730-3522

Antonio Franco Vadillo
Instituto Politécnico Nacional, México
ORCID: 0000-0002-2904-1709

Mireille Toledo Blas
Instituto Politécnico Nacional, México
ORCID: 0000-0003-3789-0898

Fabián Gómez Santiago
Instituto Politécnico Nacional, México
ORCID: 0000-0001-5699-5754

Héctor Alfredo Baptista Gonzales
Hospital Medica Sur, México
ORCID: 0000-0003-0340-0942

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.03

03.

Inteligencia artificial y paleogenómica predictiva: Incompatibilidad Rh y Kell en el colapso demográfico neandertal

Introducción

La paleoantropología reconstruye la filogenia humana mediante registros fósiles y genética, integrando disciplinas como geología para comprender el Pleistoceno. El género *Homo* evolucionó mediante divergencias y mezclas complejas desde un nodo común (*Homo antecessor*), seguidas de expansiones migratorias. Este proceso incluyó un aislamiento geográfico prolongado (~450 ka) en los linajes arcaicos de Eurasia, lo que permitió que la deriva genética fijara variantes sanguíneas únicas antes del encuentro con *Homo sapiens*. Esta hibridación con neandertales y denisovanos dejó una huella persistente en euroasiáticos y poblaciones de Oceanía, coexistiendo con otros homininos como *H. floresiensis* y *H. luzonensis* en un paisaje bionómico que determinó marcadores de inmunidad relevantes para nuestra especie actual (Figura 1).

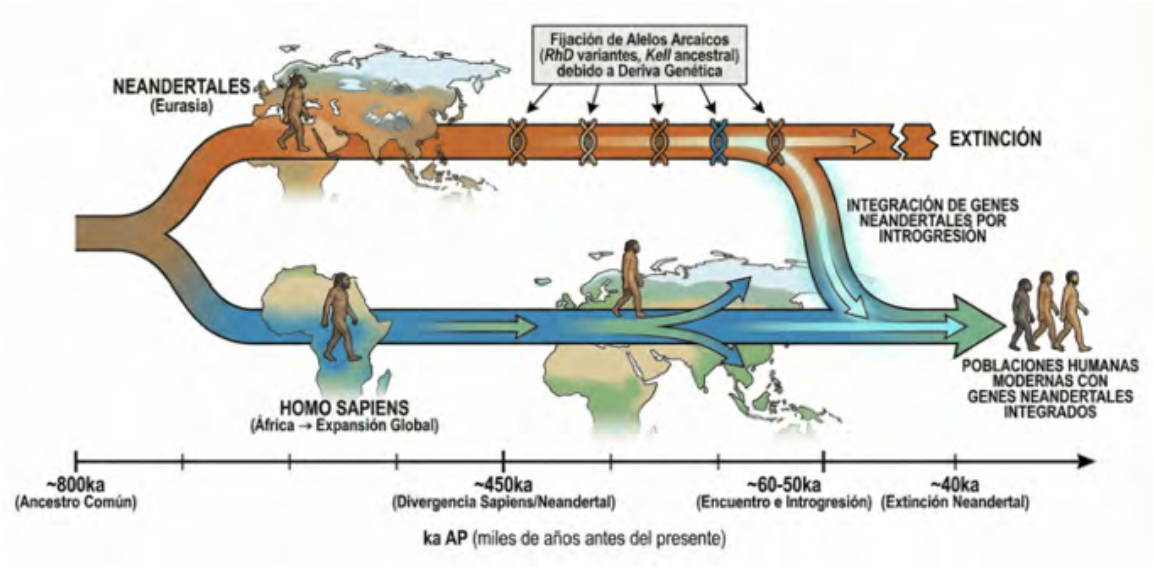
La investigación del ADN antiguo (aDNA) transitó del estudio mitocondrial hacia la secuenciación nuclear de alta resolución (Mazières et al., 2025). Este avance en técnicas ómicas facilitó una transición desde la antropología física convencional hacia una biología molecular de precisión aplicada al Paleolítico (Condemi et al., 2021). Bajo esta perspectiva, los sistemas de grupos sanguíneos emergen como descriptores polimórficos esenciales para dilucidar flujos migratorios permitiendo establecer correlaciones geográficas sobre la dispersión poblacional y mezcla génica (Cavalli-Sforza & Bodmer, 1999; Lisker, 1981; Pääbo, 2014; Kim et al., 2021). En este marco, la caracterización inmunohematológica permite reconstruir la identidad ancestral y las barreras biológicas invisibles que moldearon la trayectoria evolutiva de supervivencia de nuestra especie humana actual.

La disposición de bibliotecas genómicas de alta fidelidad, procedentes de Altai, Chagyrskaya y Vindija, facilita el estudio de aDNA en secuencias de gran cobertura para identificar SNPs que determinan fenotipos eritrocitarios. Ello posibilita explorar dimensiones

bionómicas críticas como la susceptibilidad diferencial a patógenos y la aptitud reproductiva (*fitness*) de los híbridos (Villanea et al., 2021). Asimismo, modelos estadísticos basados en la coalescencia sugieren estructuras ancestrales profundas con divergencias hace ~1500 ka y reencuentros hace ~300 ka, lo que explica por qué una fracción significativa del genoma moderno deriva de poblaciones ancestrales divergentes (Cousins et al., 2025; Scerri et al., 2018) bajo nueva perspectiva.

Figura 1

Cronología evolutiva de la divergencia del género Homo



Nota. La línea de tiempo ilustra el periodo de aislamiento euroasiático de los neandertales (~450 ka), donde la deriva genética y el aislamiento geográfico prolongado fijaron variantes sanguíneas únicas. Se destaca el posterior encuentro con Homo sapiens (~60 ka) que resultó en eventos de introgresión y la eventual extinción del linaje arcaico. Las fechas son representativas únicamente. ka (miles de años).

Desarrollo

Remediación del ADN antiguo mediante inteligencia artificial: protocolos de autenticación y genotipado de alta resolución

Los avances en las metodologías paleogenómicas han permitido la recuperación de genomas procedentes de linajes humanos arcaicos de aproximadamente 430 ka, e incluso de restos de fauna pleistocénica con una antigüedad superior a los 1000 ka. No obstante, la reconstrucción fidedigna del ADN antiguo (aDNA) enfrenta desafíos críticos como la

contaminación exógena, la fragmentación extrema de las secuencias, la ineficiencia en los procesos de biblioteca y, fundamentalmente, las lesiones químicas de carácter *post-mortem*.

El material genético antiguo exhibe una firma tafonómica altamente heterogénea resultante de una degradación cinética que induce la fragmentación aleatoria de las cadenas polinucleotídicas. Este daño altera la reconstrucción de los marcos de lectura abiertos (*Open Reading Frames,* ORF) en *loci* de interés (Borry et al., 2021). El proceso bioquímico más característico es la desaminación hidrolítica de la citosina, cuyo producto, el uracilo, es interpretado por la polimerasa como timina. Esto genera transiciones espurias de tipo C→T concentradas en los extremos 5' de las lecturas. Para neutralizar estos sesgos, la paleogenómica contemporánea ha integrado arquitecturas de Inteligencia Artificial (IA) en diversos estratos del flujo de trabajo bioinformático.

La implementación de redes neuronales convolucionales (CNN) ha resultado disruptiva al abordar el genotipado como un problema de visión artificial (Tabla 1). Frameworks de vanguardia como *DeepVariant* transforman los archivos de alineamiento (BAM) en tensores de imágenes de *pileup,* donde múltiples canales codifican propiedades de la secuencia: calidad de base, cadena de lectura, posición del fragmento y discrepancias de alineamiento. Estas representaciones son procesadas por arquitecturas como *InceptionV3*, entrenadas con genomas de referencia de alta fidelidad para discriminar con precisión entre errores inducidos por la diagénesis molecular y variantes biológicas endógenas (Poplin et al., 2018) (Figura 2).

Tabla 1

Herramientas de IA y Frameworks empleados en la remediación y autenticación de secuencias de ADN antiguo

Herramienta / Framework	Metodología de IA	Función principal en paleogenómica	Referencia
PyDamage	*Machine Learning*	Identificación automatizada de daños *post-mortem* mediante perfiles de desaminación y longitud.	(Borry et al., 2021)
DeepChem-Variant	*Deep Learning* (CNN)	Genotipado de variantes mediante el reconocimiento de imágenes de alineamiento (*pileups*).	(Bisoi et al., 2026)
AuthentiCT	Modelos bayesianos	Estimación de la contaminación mediante la distribución de sustituciones C→T en los extremos.	(Peyrégne & Peter, 2020)
mapDamage2.0	Inferencia estadística	Cuantificación de patrones de desaminación y probabilidad de daño por fragmento.	(Jónsson et al., 2013)

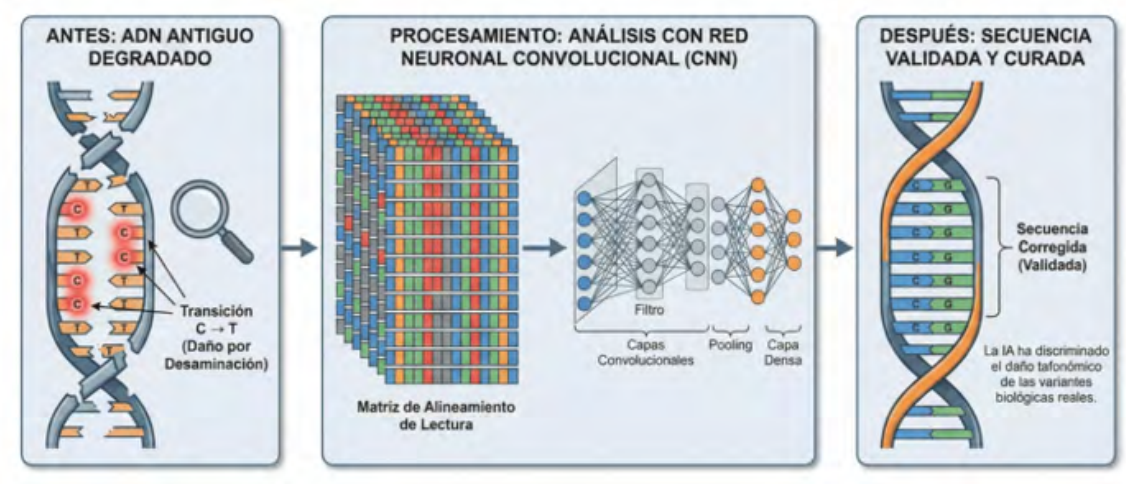
Tabla 1

Herramientas de IA y Frameworks empleados en la remediación y autenticación de secuencias de ADN antiguo

Herramienta / Framework	Metodología de IA	Función principal en paleogenómica	Referencia
PMDtools	Filtros estocásticos	Separación de lecturas con daño característico para asegurar la endogeneidad de la muestra.	(Skoglund et al., 2014)

Figura 2

Remediación de aDNA mediante inteligencia artificial



Nota. Se ilustra el procesamiento de fragmentos degradados. En el centro, la CNN analiza la matriz de alineación para discriminar el daño tafonómico de las variantes biológicas reales, lo que permite una curación precisa de la secuencia.

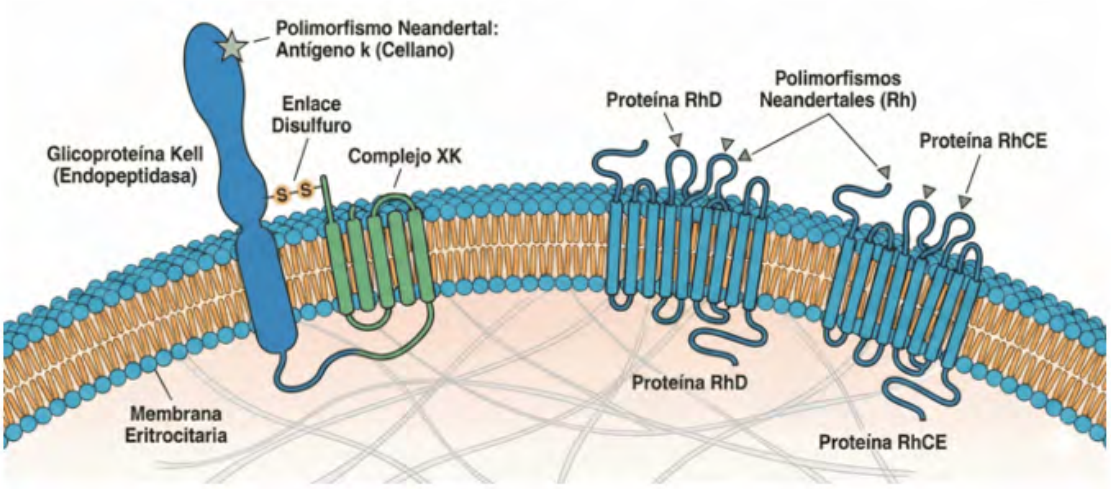
Más allá de la identificación de variantes, la IA optimiza el ensamblaje mediante aprendizaje por refuerzo, especialmente en sistemas donde la longitud promedio de los fragmentos es inferior a 50 pares de bases (bp). En regiones genómicas repetitivas o estructuralmente complejas, como los genes del sistema Rh, la ambigüedad del alineamiento es crítica. El uso de transformadores (*transformers*), capaces de analizar el contexto local y global de las lecturas, reduce significativamente el riesgo de alineamientos erróneos en parálogos similares. 2025). Esta convergencia tecnológica permite una reconstrucción virtual de haplotipos sanguíneos neandertales con una fiabilidad estadística comparable a la de contextos clínicos modernos (Chang et al., 2024).

El sistema Kell en los humanos arcaicos: un desafío de predicción fenotípica y bioquímica molecular avanzada

El sistema de grupos sanguíneos Kell (ISBT 006) está constituido por al menos 38 antígenos y es conocido por su polimorfismo y su alta inmunogenicidad, solo superados por los sistemas ABO y Rh. Su importancia radica en estar asociado a enfermedad hemolítica del feto y recién nacido y a reacciones postransfusionales. La glicoproteína Kell está codificada por el gen *KEL* en el cromosoma 7q34, compuesto por 20 exones que abarcan 21.25 kb de ADN genómico. A este gen también se le conoce como metalopeptidasa Kell, ECE3 o CD238. Funciona como una endopeptidasa de tipo II que se expresa de manera preeminente en la superficie de los eritrocitos y sus precursores eritroides (Davison et al., 2025). Los polimorfismos de un solo nucleótido son responsables de los múltiples antígenos del sistema Kell (Maheshwari & Zubair, 2025). Estructuralmente, es una proteína transmembrana de paso único conectada covalentemente mediante puentes disulfuro a la proteína XK, cuya ausencia se asocia con el síndrome de McLeod. Así, el gen *XK* funciona como un gen asociado y es necesario para la expresión del antígeno Kell. El sistema de grupo sanguíneo KX (ISBT 019) es codificado por el gen *XK*, que se encuentra en el brazo corto del cromosoma X (Xp21.1) y es responsable de la formación del antígeno K (Figura 3).

Figura 3

Diagrama Molecular de los Sistemas Kell y Rh en la membrana eritrocitaria del Pleistoceno



Nota. Representación bionómica que destaca la unión Kell-XK mediante puentes disulfuro y la estructura multipaso de las proteínas Rh. Se señalan los sitios de polimorfismo neandertal identificados como el antígeno k (Cellano).

El efecto patogénico fundamental de los anticuerpos anti-Kell es suprimir activamente la eritropoyesis fetal al interactuar con las células progenitoras en la médula ósea, lo que provoca una anemia hiporregenerativa severa (Chou & Westhoff, 2010; Ohto et al., 2023). La investigación paleogenómica ha logrado descifrar la arquitectura del locus *KEL* en homininos arcaicos. El análisis de genomas de alta cobertura ha permitido identificar a la doble dosis del alelo KEL*02, que codifica el antígeno Cellano (ISBT 006 002) (Condemi et al., 2021), es decir, las poblaciones neandertales eran Cellano (ISBT 006 002). En cuanto a la presencia del antígeno K (K1), lo que las situaba en un estado bionómico vulnerable ante la introducción de alelos divergentes por parte de otros grupos de homininos (Tabla 2).

Tabla 2

Comparativa de perfiles inmunohematológicos humanos arcaicos y modernos

Sistema sanguíneo	Neandertales (Vindija/ Altai/Chagyrskaya)	Denisovanos (Denisova 3)	Humanos modernos (No africanos)
Kell (KEL)	Homocigotos k/k (Cellano)	Homocigotos k/k	Predominio k/k (K minoritario)
Rh (RHD)	Variantes parciales / D-positivo único	D-positivo ancestral (Completo)	D-positivo / Deleción (D-negativo)
Rh (RHCE)	Combinaciones RHCEce / RHCEcE	RHCEce ancestral	Alta variabilidad (Ce, cE, ce, CE)
Genotipo RHD	RHD*DIII type4 (602G y 733C)/ RHD*DIII type4 (602G y 733C) RHD*DIII type4 (602G y 733C)/ RHD*DIII type4 (602G y 733C). RHD*DIII type4 (602G y 733C)/ RHD*DIII type4 (602G y 733C)	RHD*01/ RHD*DIII type4 (929G y sin 455C)	56 alelos de ambos genes

Tabla 2

Comparativa de perfiles inmunohematológicos humanos arcaicos y modernos

Sistema sanguíneo	Neandertales (Vindija/ Altai/Chagyrskaya)	Denisovanos (Denisova 3)	Humanos modernos (No africanos)
Genotipo RHCE	RHCE*ceEK/ *ce(48C, 744C, 787G) RHCE*ceEK/ *ce(48C, 410T, 712G) RHCE*ceEK/ *ce(48C, 410T, 800A)	RhCE*ce(48) / *ce(48C, 410T)	
Fenotipo deducido	RhD parcial	RhD positivo	RhD positivo y mecanismos moleculares predominantes para la condición RhD negativo
Estatus Genético	Baja diversidad (Alta tasa de ROH)	Diversidad moderada	Alta diversidad por expansión poblacional
Haplotipos	“Haplotipos Fantasma” (Únicos)	Perfil Rh ancestral conservado	Diversidad restringida por cuello de botella

Adicionalmente, en tres muestras de neandertales se identificó el fenotipo Js(a+b-) del sistema Kell debido a la doble dosis del alelo KEL06 (ISBT 006 006). La combinación Js(a+b-) está ausente en poblaciones europeas y asiáticas y es frecuente en poblaciones africanas (Kim et al., 2021). Por el contrario, Denisova 3 presenta el fenotipo deducido Js(a-b+) resultante de una dosis doble del alelo *KEL07* antitético (Condemi et al., 2021). Queda abierta la discusión sobre las diferencias en la incidencia del alelo *KEL06*, presente en neandertales europeos, ausente en denisovanos, así como en europeos y asiáticos modernos; pero presente en población de ascendencia africana reciente.

Mediante el uso de herramientas como AlphaFold2 y métodos de predicción a escala de biobanco (Hyvärinen et al., 2024), se ha podido inferir que ciertas sustituciones de aminoácidos en la endopeptidasa neandertal podrían modificar la accesibilidad de los epítopos a los anticuerpos maternos en contextos de aloinmunización. La técnica de imputación genotípica basada en paneles de referencia modernos, potenciada por herramientas de IA (Biagini, 2025), ha sido fundamental para cerrar las brechas en la secuencia neandertal. Estos modelos permiten inferir alelos en sitios de baja cobertura mediante la comparación de haplotipos ancestrales conservados. Curiosamente, los hallazgos indican

que los neandertales compartían ciertos alelos raros con poblaciones actuales del África subsahariana, lo que sugiere una retención de polimorfismos ancestrales que se perdieron durante el cuello de botella genético asociado con la migración OOA (Condemi et al., 2021; Mazières et al., 2025).

El sistema Rh y la complejidad estructural y parología génica en el Paleolítico superior

El RH constituye uno de los polimorfismos de mayor complejidad en el genoma humano debido a la presencia de dos genes parálogos estrechamente vinculados, *RHD* y *RHCE* (*ISBT 004*), situados en tándem en el brazo corto del cromosoma 1 (1p36.11). Estos genes, que presentan una homología de secuencia superior al 97%, son el resultado de una duplicación génica antigua y se encuentran orientados en direcciones opuestas. Esta disposición genómica facilita eventos de conversión génica, deleciones de exones y rearreglos complejos mediables por recombinación no homóloga durante la meiosis, lo que da lugar a una vasta diversidad de variantes fenotípicas de los alelos *RHD*/*RHCE*, que van desde la ausencia del gen RH, mecanismo más común del fenotipo RhD negativo, con la disminución epítopos RhD débil o diferencias en la expresión antigénica o RhD variantes, hasta la ausencia de ambos genes o Rh nulo, lo que históricamente han desafiado los algoritmos de alineamiento bioinformático estándar (Ramsey, 2025).

En el estudio de la diversidad genética de los grupos sanguíneos de 22 *Homo sapiens* y 14 neandertales de Eurasia que vivieron entre 120 ka y 20 ka antes del presente (AP). La diversidad de alelos de grupos sanguíneos en los neandertales se mantuvo sin mayores cambios, mientras que *H. sapiens* aportó a la reserva genética alelos de grupos sanguíneos actualmente más frecuentes en poblaciones no africanas, lo que sugiere que podrían haberse diferenciado justo después de la OOA, entre 70 y 45 ka. En la población siberiana, de origen paleolítico superior, se observaron alelos desconocidos que podrían ilustrar la pérdida de la herencia genética de los primeros euroasiáticos.

En el análisis comparativo entre los neandertales y denisovanos de los alelos *RHD*/*RHCE* y los fenotipos inducidos (Condemi et al., 2021), se identificaron tres alelos variantes *RHCE*ceEK (c.48G>C, c.712A>G, c.787A>G y quizá c.800T>A) en Altai Neandertal, Vindija y Chagyrskaya, y dos *RHCE*ce de tipo salvaje en Denisova-3. Cada uno de los genomas arcaicos neandertales presenta otra variante africana, *RHCE*ce, *en cis* (Tabla 2).

La implementación de IA y otras herramientas bioinformáticas permite predecir la inmunogenicidad de estas proteínas ancestrales mediante el análisis de sus estructuras proteicas secundarias y terciarias mediante simulaciones de dinámica molecular. Al modelar las sustituciones de aminoácidos identificadas en el gen RHD de neandertales, se ha revelado que su conformación molecular constituye una firma biológica de una evolución aislada en el continente euroasiático durante más de 400 ka (Velasco Cardona et al., 2025). La identificación de estos haplotipos únicos en muestras de ADN degradado ha sido posible gracias a algoritmos de imputación que comparan genomas arcaicos con bases de datos de biobancos a gran escala, lo que permite detectar segmentos de introgresión que, de otro modo, pasarían inadvertidos como ruido de la secuenciación técnica (Biagini et al., 2025).

Implicaciones evolutivas: Incompatibilidad materno-fetal, ley de Hardy-Weinberg y modelado estocástico del Pleistoceno

La caracterización bionómica de los sistemas de grupos sanguíneos Rh y Kell en las poblaciones de homininos arcaicos no constituye únicamente un ejercicio de catalogación genética, sino que fundamenta una de las hipótesis más robustas sobre el declive de los neandertales: la Enfermedad Hemolítica del Recién Nacido (EHRN) como una barrera biológica crítica y acumulativa. Este fenómeno, un trastorno inmunológico complejo mediado por la transferencia transplacentaria de anticuerpos maternos del isotipo IgG contra antígenos eritrocitarios fetales, habría alcanzado niveles de riesgo demográfico significativos tras los eventos de hibridación entre hembras neandertales y varones *Homo sapiens* o denisovanos (Condemi et al., 2021; Sullivan et al., 2017). La divergencia antigénica, producto de milenios de aislamiento geográfico en Eurasia, creó un escenario donde el encuentro genético, lejos de ser puramente beneficioso por el aporte de diversidad, introdujo incompatibilidades biológicas profundas que comprometieron la viabilidad de la descendencia híbrida en un proceso de selección purificadora de alta intensidad.

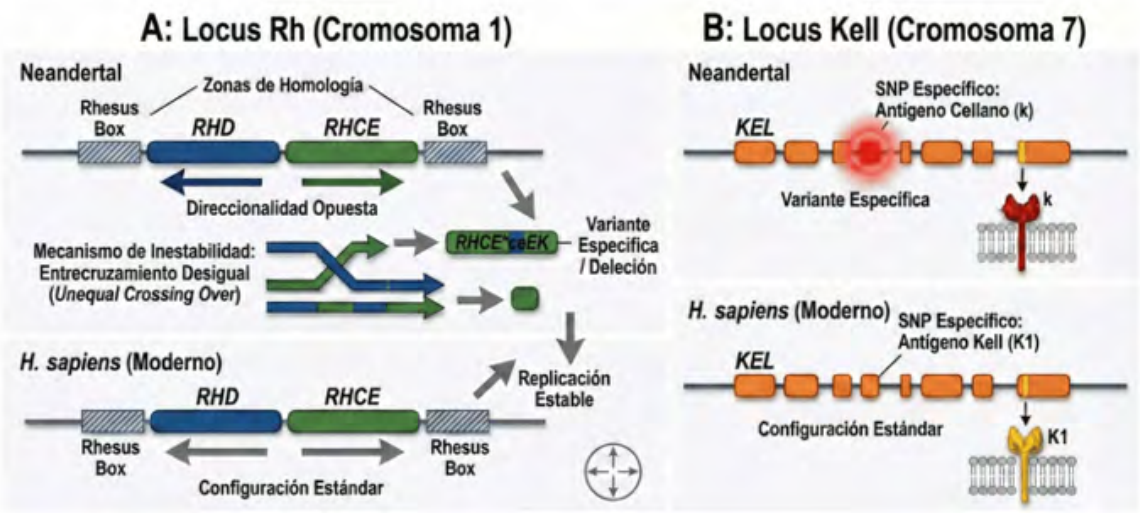
Desde la perspectiva de la genética de poblaciones, la EHRN induce una desviación sistemática y letal de la ley de Hardy-Weinberg ($p^2 + 2pq + q^2 = 1$). En un sistema en equilibrio, las frecuencias alélicas tienden a la estabilidad bajo condiciones ideales de apareamiento aleatorio y ausencia de presiones externas; sin embargo, la presencia de aloinmunización materna actúa como una fuerza de selección negativa que se ensaña específicamente con el heterocigoto o el genotipo incompatible. Hallazgos paleogenómicos en especímenes de Altai, Vindija y Chagyrskaya han confirmado la presencia del alelo variante RHCEceEK (caracterizado por las sustituciones c.48G>C, c.712A>G, c.787A>G). En un estado de homoci-

gosis, este alelo codifica un fenotipo carente del antígeno público RH:-18, una configuración extremadamente inusual en humanos modernos pero que parece haber sido prevalente en los linajes arcaicos. La arquitectura del locus RH es intrínsecamente propensa a reordenamientos complejos debido a su naturaleza de parálogos en tándem, lo que facilita eventos de conversión génica y deleciones de exones que desafían los algoritmos de alineamiento estándar (Ramsey, 2025). Esta inestabilidad estructural se muestra en la Figura 4 e ilustra cómo la disposición contrapuesta de los genes *RHD* y *RHCE* y la presencia de las "*Rhesus boxes*" predispusieron al linaje neandertal a errores de copia y deleciones recurrentes, dificultando la replicación estable observada en genomas modernos.

En los neandertales, esta configuración genética desencadenaba respuestas inmunitarias graves similares a la incompatibilidad RhD contemporánea, pero con la agravante de que el sistema Kell también presentaba una vulnerabilidad análoga debido a la fijación del alelo *KEL02*, responsable del antígeno Cellano. Esta doble presión inmunitaria sobre el feto no solo causaba anemia hemolítica severa, sino que comprometía la aptitud biológica (*fitness*) de toda la comunidad al reducir drásticamente el éxito reproductivo en los encuentros de mezcla poblacional.

Figura 4

Arquitectura genómica de los loci RH y KEL y su susceptibilidad a reordenamientos



La introgresión genómica resultante de estos encuentros entre neandertales y humanos modernos presenta un carácter profundamente paradójico que la paleogenómica actual busca descifrar con precisión molecular. Por un lado, la herencia neandertal confirió ventajas adaptativas inmediatas a los Homo sapiens que iniciaron la expansión fuera de África (OOA), proporcionando variantes genéticas optimizadas para resistir patógenos virales locales y adaptarse a regímenes de radiación ultravioleta reducidos (Zhou et al., 2023). Sin embargo, esta misma herencia incrementó la susceptibilidad a trastornos autoinmunes y respuestas inflamatorias exacerbadas en contextos modernos. Más allá de la inmunidad, la persistencia de variantes deletéreas en genes reguladores de la reproducción y del sistema inmunológico —como UNC13D, asociado a linfohistiocitosis hemofagocítica, y MOGS (manosil-oligosacárido glucosidasa), vinculado a defectos críticos de glicosilación— sugiere que los híbridos cargaban con una carga genética onerosa que limitaba su viabilidad a largo plazo. Esta dinámica de selección purificadora se evidencia con rigor científico mediante el uso de algoritmos de coalescencia y modelos de estructura ancestral compartida (Cousins et al., 2025). Estos modelos, basados en Modelos Ocultos de Markov (HMM), funcionan como herramientas estadísticas de alta precisión que asumen que el genoma contemporáneo es un mosaico de ancestros diversos. Los HMM permiten analizar el sistema como un proceso estocástico donde los estados ocultos (la ascendencia neandertal o sapiens en un segmento genómico específico) emiten datos visibles (los alelos secuenciados).

Esta metodología ha sido fundamental para identificar los denominados “desiertos de introgresión”, un término acuñado originalmente por Sankararaman et al. (2014) para describir regiones del genoma humano moderno donde la frecuencia de ascendencia arcaica es nula o extremadamente reducida. La existencia de estos desiertos implica que cualquier variante arcaica introducida originalmente en esas coordenadas genómicas fue eliminada con celeridad mediante una selección purificadora de alta intensidad, lo que sugiere que dichos segmentos contenían incompatibilidades biológicas profundas que comprometían la fertilidad de sus portadores. No es casualidad que estos desiertos coincidan casi quirúrgicamente con genes críticos para la salud gestacional y el desarrollo neuronal. Como se ilustra en el paisaje genómico de la Figura 5, mientras que algunas regiones muestran picos de introgresión adaptativa (donde los genes neandertales fueron conservados por su valor biológico), los valles o desiertos representan las zonas donde la incompatibilidad era tan profunda que los portadores simplemente no lograron legar su material genético a las siguientes generaciones (Kendall et al., 2025). La afectación de la fertilidad femenina representa, quizás, el punto más crítico de esta carga. La identificación de variantes perjudiciales

en 55 genes controladores de funciones ováricas y uterinas, tales como AKAP1, BUB1B, ESR1 y WNT4, junto con desequilibrios de ligamiento en *loci* metabólicos como CYP2C8 y CYP2C9, refuerza la tesis de que las restricciones biológicas internas desempeñaron un papel determinante en la dinámica demográfica del Paleolítico (Greer et al., 2021).

Figura 5

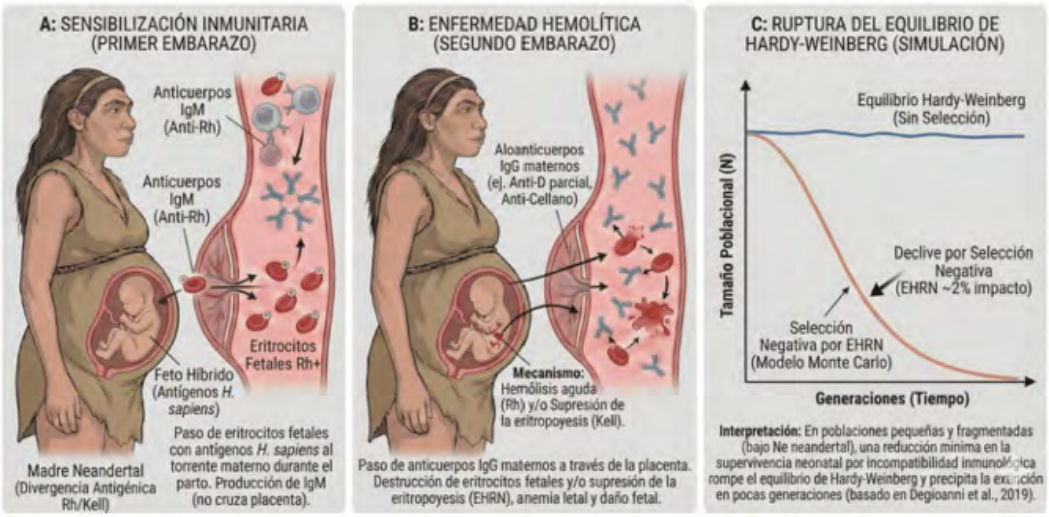
Paisaje genómico de la introgresión neandertal y desiertos de selección negativa

Para comprender la magnitud real de este impacto demográfico, es imperativo trascender los modelos deterministas tradicionales y recurrir a los fundamentos de las simulaciones estocásticas y el método de Monte Carlo. Los modelos deterministas asumen resultados fijos basados en parámetros iniciales constantes, pero la biología del Pleistoceno estaba sujeta a la incertidumbre inherente de poblaciones pequeñas, eventos climáticos extremos y una deriva genética acentuada. Las simulaciones estocásticas incorporan variables aleatorias para reflejar esta realidad multicausal, entendiendo que el destino de un alelo neandertal no dependía únicamente de su valor adaptativo intrínseco, sino también del azar genético en poblaciones con un tamaño efectivo (N_e) extremadamente reducido. Mediante el uso de algoritmos de Monte Carlo, se ejecutan millones de iteraciones de historias de vida virtuales donde el sistema computacional asigna probabilidades de ocurrencia aleatorias para determinar variables críticas de salud reproductiva: ¿heredó el feto el alelo incompatible?, ¿fue esta la segunda gestación de la madre, facilitando la sensibilización inmunitaria previa?, ¿resultó el embarazo en una pérdida reproductiva o en un individuo con compromiso sistémico severo?

Los resultados de estos modelos demuestran que un decremento demográfico secular en el éxito reproductivo, incluso de magnitudes tan reducidas como un 1% o 2% acumulado durante cientos de generaciones, es suficiente para precipitar un declive poblacional irreversible (Degioanni et al., 2019). En una población con baja densidad demográfica como la neandertal, esta erosión constante de la aptitud biológica actuó como un lastre bionómico que ninguna ventaja adaptativa colateral pudo compensar. Este complejo mecanismo fisiopatológico y su traducción en curvas de colapso exponencial se sintetizan en el modelo de la Figura 6, donde se observa la correlación directa entre la sensibilización inmunitaria placentaria (Paneles A y B) y el alejamiento drástico del equilibrio poblacional teórico (Panel C). No se trataba exclusivamente de una competencia por recursos ecológicos, sino de una incapacidad para mantener una tasa de reemplazo poblacional positiva frente a una especie invasora con sistemas inmunitarios y reproductivos más generalistas.

Figura 6

Mecanismo fisiopatológico y estadístico del colapso neandertal por EHRN



La síntesis de los datos paleogenómicos e inmunohematológicos permite trazar un hilo conductor entre la arquitectura cromosómica y el colapso demográfico de los linajes arcaicos. La inestabilidad estructural intrínseca de los genes parálogos *RHD* y *RHCE*, cuya configuración en tándem favorece la recombinación no homóloga y errores de copia (Figura 4), no representó un simple polimorfismo, sino una vulnerabilidad biológica fundamental que predispuso al linaje neandertal a una fragilidad reproductiva acumulativa que desafió los mecanismos de reparación celular. Este fenómeno se tradujo en una selección purificadora de tal magnitud que sus huellas permanecen grabadas en el genoma humano moderno

bajo la forma de 'desiertos de introgresión' (Figura 5); regiones donde la carga de variantes deletéreas era tan onerosa que la selección natural eliminó sistemáticamente cualquier rastro de ascendencia arcaica para preservar la viabilidad de nuestra especie. Al cuantificar mediante modelos estocásticos el impacto de las micro-interacciones moleculares en la placenta (Figura 6), se hace evidente que el transporte activo de inmunoglobulinas IgG maternas contra antígenos fetales divergentes, como el RH18 o el cellano, desencadenó crisis de aloinmunización de una celeridad cinética implacable, generando una memoria inmunológica que penalizaba sistemáticamente las gestaciones subsecuentes. Por tanto, el ocaso de los neandertales no se dirimió exclusivamente en el ámbito de la competencia por recursos o nichos ecológicos, sino en la escala microscópica de la interfaz hemato-placentaria. En este espacio bionómico, la inmunidad y la genética dictaron el final de un linaje y el éxito adaptativo del nuestro, convirtiendo el estudio detallado de estos picos y valles de introgresión en una herramienta diagnóstica indispensable para comprender la arquitectura biológica de la salud y la enfermedad en la humanidad contemporánea.

El análisis se vuelve aún más complejo al integrar el método basado en la coalescencia denominado COBRAA (*Coalescent-Based Record of Ancestry and Admixture*). Mediante este enfoque, se ha inferido una estructura ancestral profunda en nuestros antepasados, sugiriendo una división en la ascendencia humana hace aproximadamente 1500 ka, seguida de un evento único de mezcla poblacional hacia los 300 ka (Cousins et al., 2025). Esta evidencia, que es independiente de la secuenciación directa de genomas arcaicos, sugiere que los neandertales y denisovanos divergieron de la línea principal (Población A) de manera más reciente que la línea secundaria (Población B). El porcentaje de mezcla de aproximadamente el 20% identificado en estos modelos es significativamente mayor que la proporción de mezcla neandertal o denisovana presente en las poblaciones no africanas actuales (~2%), lo que sugiere que hubo un proceso masivo de cribado genético tras los encuentros iniciales. Este modelo de estructura ancestral plantea preguntas fundamentales sobre la relación entre los linajes A y B y los homininos previamente identificados por la arqueología, como *H. erectus* o *H. heidelbergensis*. El dramático cuello de botella observado en la Población A tras el evento de mezcla podría estar directamente relacionado con las incompatibilidades inmunohematológicas descritas en este estudio, actuando como un filtro biológico que determinó la viabilidad de los linajes resultantes.

La profundidad de estos hallazgos sugiere que la extinción de los homininos arcaicos debe ser reinterpretada no solo como un fenómeno multicausal, sino como un proceso donde

la biología molecular dictó las reglas de la supervivencia demográfica. La integración de la inmunohematología clínica con la paleogenómica de alta resolución permite una relectura del registro fósil desde una perspectiva fisiológica y patológica. La pérdida sistemática de descendencia debido a la EHRN en parejas mixtas no fue un evento anecdótico, sino un catalizador biológico de la extinción definitiva. En una población ya mermada por factores ambientales extremos, la introducción de incompatibilidades en los sistemas Rh y Kell (especialmente el antígeno Cellano) actuó como un mecanismo de selección negativa que redujo la aptitud biológica de los híbridos y aisló reproductivamente a los grupos arcaicos.

Conclusiones y perspectivas futuras

La sinergia entre inteligencia artificial y paleogenómica ha permitido trascender la catalogación fósil para reconstruir la fisiología molecular de especies extintas. El análisis de genomas de alta fidelidad (Prüfer et al., 2014, 2017), refinado mediante arquitecturas de aprendizaje profundo (Bisoi et al., 2025), revela una complejidad bionómica única en los sistemas Rh y Kell neandertales, cuya huella persiste en la variabilidad humana actual debido a la introgresión adaptativa (Condemi et al., 2021).

Desde una perspectiva microevolutiva, la enfermedad hemolítica del recién nacido (EHRN) se identifica como un factor de selección negativa crítico. Las crisis de aloinmunización materna actuaron como barreras reproductivas invisibles que alteraron sistemáticamente el equilibrio de Hardy-Weinberg, sugiriendo que el colapso de los homininos arcaicos no fue un evento meramente multicausal, sino impulsado por restricciones biológicas internas inalienables detectables mediante modelos de simulación estocástica (Degioanni et al., 2019).

Finalmente, la transformación digital mediante herramientas como *PyDamage* y *DeepChem-Variant* ha dotado a la bioquímica clínica de una dimensión retrospectiva esencial para comprender las raíces evolutivas de patologías inmunitarias modernas. El futuro de la investigación reside en la integración de *Big Data* y paleogenómica predictiva para revelar la arquitectura biológica de nuestra especie, convirtiendo el ADN antiguo en una narrativa dinámica sobre los desafíos inmunológicos que moldearon el destino de la humanidad.

Referencias

Biagini, S. A., Becelaere, S., Aerden, M., Jatsenko, T., Hannes, L., Van Damme, P., Breckpot, J., Devriendt, K., Thienpont, B., Vermeesch, J. R., Cleynen, I., & Kivisild, T. (2025). Genotype imputation from low-coverage data for medical and population genetic analyses. *Genome research*, *35*(9), 1929–1941. https://doi.org/10.1101/gr.280175.124

Bisoi, A. V., Jose, S., Siguenza, S., & Ramsundar, B. (2025). DeepChem-Variant: A modular open source framework for genomic variant calling. *arXiv*. https://doi.org/10.48550/arXiv.2411.11513

Borry, M., Hübner, A., & Rohrlach, A. B. (2021). PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA de novo assembly. *Bioinformatics*, *37*(10), 1452–1454. https://doi.org/10.1093/bioinformatics/btab186

Cavalli-Sforza, L. L., & Bodmer, W. F. (1999). *The Genetics of Human Populations*. Dover Publications.

Chang, T.-C., Yu, J., Wang, Z., Hankins, J. S., Weiss, M. J., Wu, G., Westhoff, C. M., Chou, S. T., & Zheng, Y. (2024). Machine learning to optimize automated RH genotyping using whole-exome sequencing data. *Blood Advances*, *8*(11), 2651–2659. https://doi.org/10.1182/bloodadvances.2023012371

Condemi, S., Mazières, S., Faux, P., Costedoat, C., Ruiz-Linares, A., Bailly, P., & Chiaroni, J. (2021). Blood groups of Neandertals and Denisova decrypted. *PLOS ONE*, *16*(7), e0254175. https://doi.org/10.1371/journal.pone.0254175

Chou, S. T., & Westhoff, C. M. (2010). El sistema Rh y los sistemas de grupos sanguíneos RhAG. *Immunohematology*, 26, 178–186.

Cousins, T., Scally, A., & Durbin, R. (2025). A structured coalescent model reveals deep ancestral structure shared by all modern humans. *Nature Genetics*, *57*(4), 856–864. https://doi.org/10.1038/s41588-025-02117-1

Davison, C. L., & Denomme, G. A. (2025). Kell and Kx blood group systems: an update. *Immunohematology*, *41*(3), 90–99. https://doi.org/10.2478/immunohematology-2025-014

Degioanni, A., Bonenfant, C., Cabut, S., & Condemi, S. (2019). Living on the edge: Was demographic weakness the cause of Neanderthal demise? *PLOS ONE*, *14*(5), e0216742. https://doi.org/10.1371/journal.pone.0216742

Greer, C., Bhakta, H., Ghanem, L., Refai, F., Linn, E., & Avella, M. (2021). Deleterious variants in genes regulating mammalian reproduction in Neanderthals, Denisovans and extant humans. *Human Reproduction*, *36*(3), 734–755. https://doi.org/10.1093/humrep/deaa347

Hyvärinen, K., Haimila, K., Moslemi, C., Biobank, B. S., Olsson, M. L., Ostrowski, S. R., et al. (2024). A machine-learning method for biobank-scale genetic prediction of blood group antigens. *PLOS Computational Biology*, *20*(3), e1011977. https://doi.org/10.1371/journal.pcbi.1011977

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* (Oxford, England), *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kendall, C., Nooranikhojasteh, A., & Parra, E. J. (2025). Archaic adaptive introgression in modern human reproductive genes. *Communications Biology*, 8, 1365. https://doi.org/10.1038/s42003-025-08682-9

Kim, T. Y., Yu, H., Phan, M. T., Jang, J. H., & Cho, D. (2021). Application of Blood Group Genotyping by Next-Generation Sequencing in Various Immunohaematology Cases. *Transfusion medicine and hemotherapy*, *49*(2), 88–96. https://doi.org/10.1159/000517565

Lisker, R. (1981). *Estructura genética de la población mexicana.* Salvat Editores

Maheshwari, A., Zubair, M., & Maheshwari, A. (2025). *Kidd blood group system*. StatPearls Publishing.

Mazières, S., Condemi, S., El Nemer, W., & Chiaroni, J. (2025). Rapid change in red cell blood group systems after the main Out of Africa of Homo sapiens. *Scientific Reports*, *15*(1), 1597. https://doi.org/10.1038/s41598-024-83023-0

Ohto, H., Ito, S., Srivastava, K., Ogiyama, Y., Uchikawa, M., Nollet, K. E., & Flegel, W. A. (2023). Asian-type DEL (RHD*DEL1) with an allo-anti-D: A paradoxical observation in a healthy multiparous woman. *Transfusion*, *63*(8), 1601–1611. https://doi.org/10.1111/trf.17465

Pääbo, S. (2014). *Neanderthal Man: In Search of Lost Genomes*. Basic Books.

Peyrégne, S., & Peter, B. M. (2020). AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. *Genome Biology*, 21, 246. https://doi.org/10.1186/s13059-020-02123-y

Poplin, R. et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, *36*(10), 983–987. https://doi.org/10.1038/nbt.4235

Prüfer, K., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature, 505(7481), 43–49. https://doi.org/10.1038/nature12886

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., Reher, D., Hopfe, C., Nagel, S., Maricic, T., Fu, Q., Theunert, C., Rogers, R., Skoglund, P., Chintalapati, M., Dannemann, M., … Pääbo, S. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* (New York, N.Y.), *358*(6363), 655–658. https://doi.org/10.1126/science.aao1887

Ramsey G. (2025). The Rh blood group system: RHD update. *Immunohematology*, *41*(2), 31–48. https://doi.org/10.2478/immunohematology-2025-007

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., y Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*(7492), 354–357. https://doi.org/10.1038/nature12961

Scerri, E. M., Thomas, M. G., Runge, K., et al. (2018). Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends in Ecology & Evolution*, *33*(8), 582–594. https://doi.org/10.1016/j.tree.2018.05.005

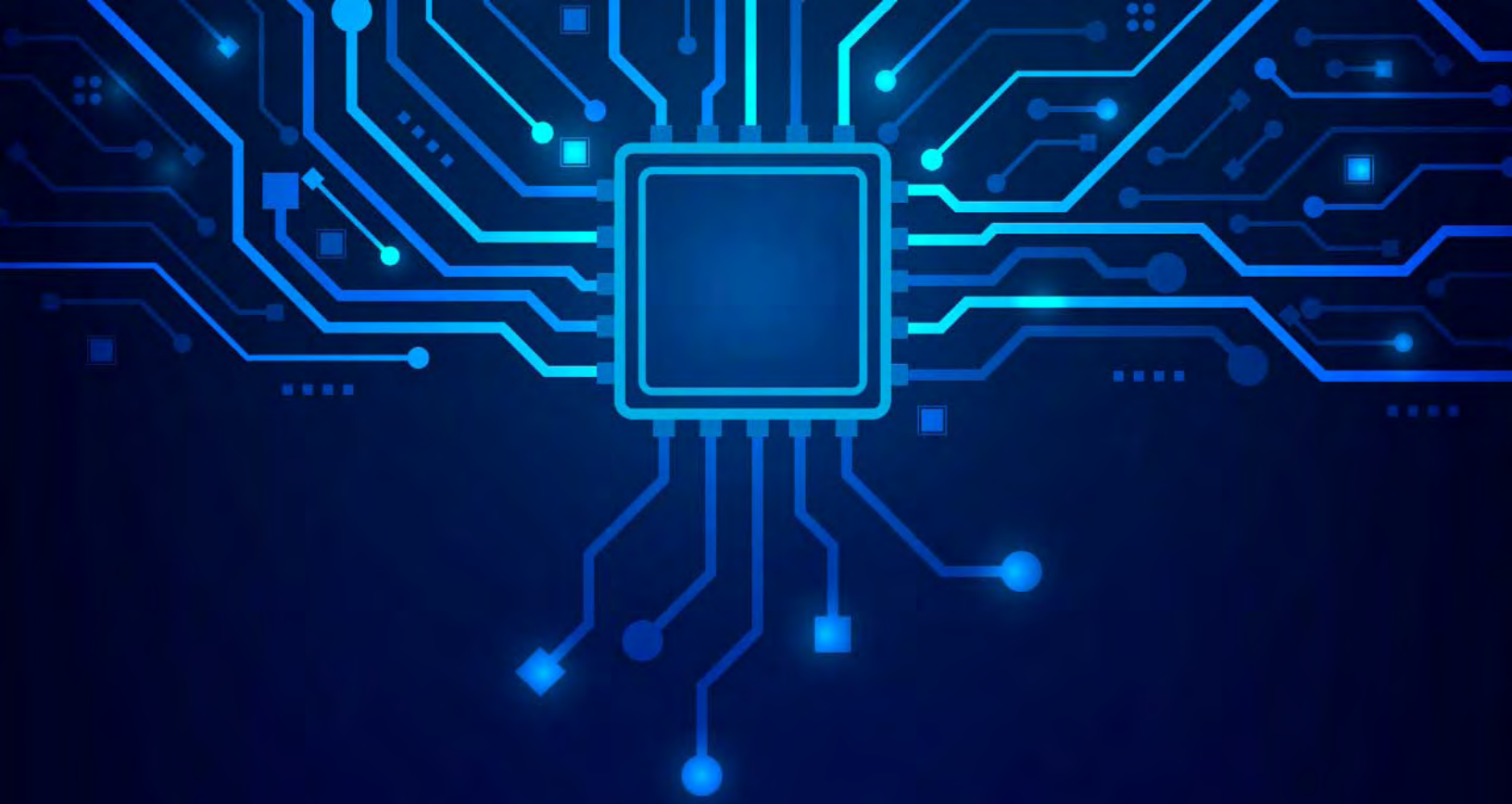
Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating ancient DNA from modern contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences*, *111*(6), 2229–2234. https://doi.org/10.1073/pnas.1318934111

Sullivan, A. P., de Manuel, M., Marques-Bonet, T., & Perry, G. H. (2017). An evolutionary medicine perspective on Neandertal extinction. *Journal of human evolution*, 108, 62–71. https://doi.org/10.1016/j.jhevol.2017.03.004

Velasco Cardona, D. C., Cardona-Acevedo, S., Valencia-Arias, A., & Valencia, J. (2025). Artificial intelligence techniques in blood banks — a systematic review of predictive innovations. *Acta Haematologica Polonica*, *56*(4), 258–268. https://doi.org/10.5603/ahp.104427

Villanea, F. A., Huerta-Sanchez, E., & Fox, K. (2021). ABO genetic variation in Neanderthals and Denisovans. *Molecular Biology and Evolution*, *38*(8), 3373–3382 . https://doi.org/10.1093/molbev/msab109

Zhou, Z., Swagemakers, S. M. A., Lourens, M. S., Suratannon, N., van der Spek, P. J., & Dalm, V. A. S. H. (2023). Did variants in inborn errors of immunity genes contribute to the extinction of Neanderthals? *Asian Pacific Journal of Allergy and Immunology*, *41*(4), 422–434. https://doi.org/10.12932/AP-251022-1489



04.

Justicia algorítmica y gobernanza ética ante los sesgos de la inteligencia artificial

Alma Delia Otero Escobar
Universidad Veracruzana, México
ORCID: 0000-0001-9266-6587

Cecilia Esperanza Ostos Cruz
Universidad Veracruzana, México
ORCID: 0009-0003-9118-9813

Elsa Suárez Jasso
Universidad Veracruzana, México
ORCID: 0000-0002-7341-1068

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.04

04.

Justicia algorítmica y gobernanza ética ante los sesgos de la inteligencia artificial

Introducción

El desarrollo acelerado de la inteligencia artificial (IA) ha transformado múltiples dimensiones de la vida moderna, desde los procesos de contratación y la gestión empresarial hasta la administración de justicia y los sistemas educativos. Sin embargo, este avance tecnológico también ha evidenciado problemáticas éticas, sociales y jurídicas derivadas del uso de algoritmos que aprenden del comportamiento humano y que, en consecuencia, pueden reproducir prejuicios históricos relacionados con la raza, el género o la condición socioeconómica (Crawford, 2021).

En los últimos años, diversos estudios han demostrado que los sistemas de IA no son completamente neutrales ni objetivos, ya que su desempeño depende de los datos con los que fueron entrenados. Cuando los conjuntos de datos contienen patrones de discriminación o exclusión, los algoritmos pueden amplificar dichas desigualdades y propiciar toma de decisiones sesgadas en ámbitos tan sensibles como la evaluación judicial, la seguridad pública, el reclutamiento laboral o la concesión de créditos financieros (O'Neil, 2016; Buolamwini & Gebru, 2018). Casos como el sistema COMPAS en Estados Unidos —utilizado para determinar el riesgo de reincidencia en procesos judiciales— o los programas de reconocimiento facial de grandes corporaciones tecnológicas que muestran menor precisión con personas afrodescendientes, son ejemplos paradigmáticos de cómo la automatización puede perpetuar el racismo estructural en entornos digitales (Angwin et al., 2016; Raji & Buolamwini, 2019).

Es así como surge la necesidad de abordar el fenómeno del racismo algorítmico desde una perspectiva integral que considere dimensiones éticas, legales y sociales. La justicia algorítmica, entendida como la capacidad de los sistemas automatizados para actuar de forma equitativa, transparente y responsable, se convierte en un desafío urgente para la gobernanza tecnológica contemporánea (Binns, 2018). Tal como señala Feenberg (2020), la tecnología no es un ente neutral, sino una construcción social que refleja los valores y

estructuras de poder de las sociedades que la producen. Por tanto, analizar la IA desde la óptica de la justicia social implica reconocer que las decisiones automatizadas pueden afectar derechos fundamentales, como la igualdad de trato y la no discriminación (Noble, 2018; Crawford, 2021).

La justificación de este estudio se fundamenta en la necesidad de comprender cómo los sesgos raciales y sociales presentes en los algoritmos de IA pueden generar consecuencias adversas en la vida de las personas. A diferencia de los errores humanos, los sesgos algorítmicos se reproducen a gran escala y con aparente objetividad, lo que dificulta su detección y corrección (Mehrabi et al., 2021). De ahí la importancia de promover una gobernanza ética de la IA que incorpore principios de transparencia, rendición de cuentas y participación inclusiva en su diseño y aplicación (Jobin et al., 2019; Floridi et al., 2020).

Mientras en Estados Unidos se ha consolidado una discusión institucional y normativa sobre los sesgos algorítmicos, en México la investigación sobre los impactos éticos y sociales de la IA sigue siendo incipiente. Aunque se han establecido políticas digitales orientadas a la innovación tecnológica, persiste un vacío normativo respecto al uso responsable de algoritmos y la protección de los derechos humanos en entornos automatizados (OECD, 2023; Secretaría de Economía, 2022). Esta situación contrasta con la experiencia de otros países, como Estados Unidos, donde los debates sobre la regulación de la IA y la prevención de la discriminación tecnológica son más amplios y sistemáticos (European Commission, 2021).

Por lo tanto, este artículo tiene como propósito analizar las implicaciones éticas, legales y sociales de los sesgos algorítmicos en los sistemas de inteligencia artificial, comparando experiencias internacionales y nacionales, y proponiendo estrategias de mitigación que contribuyan a una gobernanza más justa, inclusiva y responsable del uso de la tecnología.

En este contexto, el objetivo general del estudio es analizar las implicaciones éticas, legales y sociales de los sesgos algorítmicos en los sistemas de inteligencia artificial, comparando los contextos de México y Estados Unidos. Los objetivos específicos son: 1) identificar los principales tipos de sesgos algorítmicos y sus manifestaciones; 2) examinar los marcos normativos y éticos internacionales que orientan la gobernanza de la IA; y 3) proponer estrategias de gobernanza ética que promuevan una IA justa e inclusiva.

Método de investigación

Este estudio se desarrolló bajo un enfoque cualitativo de tipo documental, cuyo propósito fue analizar de manera crítica la literatura académica, normativa e institucional sobre los sesgos algorítmicos y su relación con la justicia y la gobernanza ética de la IA. La elección de un diseño documental responde a la necesidad de examinar interpretaciones teóricas y marcos normativos en torno a un fenómeno reciente y multidimensional que involucra tanto a la tecnología como a los derechos humanos (Merriam & Tisdell, 2016; Flick, 2018).

El proceso metodológico se estructuró en tres fases: 1) Búsqueda y recolección de información, 2) Selección y organización de fuentes relevantes, y 3) Análisis crítico y síntesis temática.

Durante la primera fase, se realizó una búsqueda sistemática en bases de datos académicas y repositorios institucionales que concentran literatura científica revisada por pares, tales como *Scopus*, *Web of Science*, *ERIC*, *SpringerLink*, *ScienceDirect*, *RedALyC* y *SciELO*, así como documentos técnicos de organismos internacionales (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], Organización para la Cooperación y el Desarrollo Económicos [OCDE], Comisión Europea, Naciones Unidas).

Para asegurar la pertinencia, se aplicaron criterios de inclusión y exclusión. Se incluyeron únicamente publicaciones entre 2018 y 2025, escritas en español o inglés, que abordaran al menos uno de los siguientes ejes temáticos: Ética de la inteligencia artificial, Sesgo algorítmico y discriminación automatizada, Gobernanza y regulación de la IA y Justicia algorítmica y derechos humanos digitales.

Se excluyeron documentos sin respaldo metodológico, notas periodísticas sin revisión técnica y publicaciones con enfoques exclusivamente técnicos, sin reflexión ética o legal.

Con el propósito de ofrecer una visión general y transparente del procedimiento metodológico seguido en este estudio, se elaboró un resumen del proceso de búsqueda, selección y análisis de las fuentes consultadas. La Tabla 1 presenta las principales fases, bases de datos utilizadas, criterios de inclusión y exclusión, así como el número de documentos seleccionados en cada etapa.

Tabla 1

Resumen del proceso metodológico del estudio documental

Fase	Actividad principal	Bases y fuentes consultadas	Criterios de inclusión	Criterios de exclusión	Documentos seleccionados
1	Búsqueda sistemática	*Scopus*, *Web of* Science, *ERIC*, *SpringerLink*, *RedALyC*, *SciELO*, UNESCO, OECD, European Commission	Publicaciones 2018–2025; idioma español o inglés; revisión por pares; pertinencia con IA, ética, sesgo y gobernanza	Documentos sin respaldo metodológico o sin enfoque ético/jurídico	85
2	Selección y depuración	Filtros por relevancia y citaciones	Coherencia temática y contribución teórica	Repetidos o sin acceso al texto completo	52
3	Análisis y síntesis	Lectura analítica y codificación	Relevancia para categorías teóricas del estudio	Enfoques exclusivamente técnicos	40

Las búsquedas se realizaron mediante operadores booleanos y truncamientos para refinar los resultados. Entre los términos utilizados estuvieron: “algorithmic bias”, “artificial intelligence AND ethics”, “AI governance”, “algorithmic justice”, “discrimination AND machine learning”, “racismo algorítmico”, “ética digital”, y “gobernanza tecnológica”.

Los resultados iniciales se filtraron por relevancia temática y número de citaciones. Posteriormente, se procedió a la lectura analítica y codificación temática (Saldaña, 2021), clasificando las fuentes en cuatro categorías analíticas:

- Fundamentos conceptuales de la IA y los sesgos algorítmicos.
- Manifestaciones del racismo algorítmico.
- Principios éticos y marcos de gobernanza.
- Propuestas de mitigación y justicia tecnológica.

El análisis siguió una lógica deductivo-inductiva, es decir, partiendo de conceptos teóricos clave se identificaron patrones y tensiones recurrentes en las fuentes. Esta técnica permitió contrastar perspectivas y construir una visión integral del problema.

En total, se revisaron 40 documentos que cumplieron con los criterios de inclusión definidos. El análisis se desarrolló bajo un enfoque temático y categorial, siguiendo la lógica

propuesta por Saldaña (2021) para la codificación cualitativa. Se identificaron patrones recurrentes, tensiones conceptuales y propuestas normativas dentro de las cuatro categorías analíticas establecidas: fundamentos conceptuales de la IA, manifestaciones del racismo algorítmico, principios éticos y marcos de gobernanza, y estrategias de mitigación. Este procedimiento permitió una comprensión comparativa y crítica de los principales debates contemporáneos sobre la justicia algorítmica.

La revisión se apoyó en un enfoque crítico e interdisciplinario, integrando aportes de la filosofía de la tecnología, el derecho digital, la sociología del conocimiento y la ciencia de datos. Este enfoque reconoce que los algoritmos no son herramientas neutrales, sino artefactos sociales que reflejan relaciones de poder, decisiones de diseño y prioridades económicas (Feenberg, 2020; Crawford, 2021).

Asimismo, se aplicó el método de triangulación conceptual, combinando fuentes teóricas, estudios empíricos y políticas públicas. Esto permitió identificar convergencias entre principios éticos (beneficencia, justicia, explicabilidad) y lineamientos normativos internacionales como los de la UNESCO (2021) y la European Commission (2021).

En cuanto a la validez y confiabilidad, para fortalecer la primera, se utilizaron estrategias de revisión cruzada entre documentos académicos y fuentes institucionales, verificando la consistencia conceptual entre distintos autores. La confiabilidad se aseguró mediante la documentación del proceso de búsqueda y el uso de citas directas con *digital object identifier* (DOI) verificable. La validez teórica se reforzó mediante la comparación entre modelos internacionales de gobernanza algorítmica (Unión Europea, Estados Unidos y México) y la aplicación de un análisis crítico orientado a los principios de equidad, responsabilidad y derechos humanos.

Esta metodología documental permitió articular un marco comprensivo de los debates contemporáneos sobre justicia algorítmica y ética digital. A través del análisis de fuentes recientes y contrastadas, se logra no solo identificar los desafíos de la IA contemporánea, sino también proponer caminos hacia una gobernanza tecnológica más transparente, inclusiva y responsable.

Entre las principales limitaciones metodológicas, se reconoce la posible presencia de sesgo de idioma, dado que la revisión se centró en publicaciones en inglés y español, excluyendo fuentes en otros idiomas. Asimismo, la disponibilidad de documentos com-

pletos en acceso abierto restringió el universo total de análisis. No obstante, la diversidad de bases consultadas y la triangulación conceptual contribuyeron a mitigar estos sesgos y fortalecer la validez del estudio.

Marco teórico

En términos generales, la IA se refiere a sistemas computacionales capaces de realizar tareas que, si fuesen realizadas por humanos, requerirían inteligencia, por ejemplo: percepción, aprendizaje, razonamiento, predicción y toma de decisiones. Más allá de definiciones instrumentales, la IA contemporánea se sustenta principalmente en técnicas de aprendizaje automático (*machine learning*) y aprendizaje profundo (*deep learning*) que aprenden patrones estadísticos a partir de datos para producir inferencias (Russell & Norvig, 2021; Mitchell, 1997). No obstante, la IA no es meramente técnica: es un sistema sociotécnico cuyo rendimiento y consecuencias dependen de datos, modelos, infraestructuras y contextos institucionales, legales y culturales (Crawford, 2021; Selbst et al., 2019). De ahí que el análisis de su impacto requiera integrar dimensiones éticas, jurídicas y sociales.

El sesgo algorítmico puede describirse como un resultado desigual generado por un sistema automatizado que afecta de manera injusta a ciertos grupos. Más que un error aislado, implica patrones estructurales que surgen desde la selección de datos hasta el uso del modelo, como han advertido Barocas y Selbst (2016), así como Friedman y Nissenbaum (1996). Puede emerger en distintas fases del ciclo de vida del sistema:

- Sesgo en los datos (*sampling*, *measurement*, *label bias*): conjuntos de datos no representativos o con etiquetas que codifican estereotipos/errores históricos (Gebru et al., 2021).
- Sesgo en el modelado: elecciones de variables, objetivos/funciones de pérdida, regularización o umbrales que optimizan métricas globales a costa de subpoblaciones (Dwork et al., 2012; Kleinberg et al., 2017).
- Sesgo en la implementación: desajustes entre el entorno de diseño y el de uso, o decisiones de despliegue que amplifican inequidades (Selbst et al., 2019).

En visión computacional y clasificación de género/rostro, se han documentado disparidades de error sustantivas por tono de piel y género, evidenciando que la representación desigual en los datos produce rendimientos inequitativos (Buolamwini & Gebru, 2018; Raji

& Buolamwini, 2019). La literatura reciente también propone instrumentos de transparencia como *model cards* (Mitchell et al., 2019) y *datasheets for datasets* (Gebru et al., 2021) para hacer explícitos *para qué* y *para quién* fueron diseñados datos y modelos, mitigando sesgos inadvertidos.

Desde el punto de vista normativo, diferentes nociones de equidad algorítmica—demographic parity, equalized odds, predictive parity, calibration—son mutuamente incompatibles en general, por lo que elegir una u otra implica juicios de valor y compensaciones (trade-offs) dependientes del contexto (Kleinberg et al., 2017; Binns, 2018). En consecuencia, la equidad no puede tratarse como un parámetro puramente técnico: requiere deliberación pública y gobernanza.

Como advierten Kearns y Roth el problema no es el algoritmo en sí, sino la falta de correspondencia entre las entradas al algoritmo y el mundo real, dado que el sesgo está incrustado en los datos. No cabe esperar que el algoritmo sea capaz de descubrirlo y corregirlo por sí mismo. *Mutatis muntandis*, se trataría de la constatación de una de las máximas más citadas de la ciencia informática: *garbage in, garbage out* (*si entra basura, sale basura*); en otras palabras, un sistema solo puede aprender a partir de los datos disponibles; si estos reflejan desigualdades o errores, el modelo inevitablemente los incorporará en sus predicciones (Llano, 2024, p. 74-75)

Ahora bien, el racismo puede entenderse como un sistema de creencias, prácticas e instituciones que asignan valor diferencial y jerarquías a grupos racializados, produciendo desigualdades estructurales (Omi & Winant, 2014). En entornos digitales, el racismo algorítmico describe la reproducción y, a veces, intensificación de esas desigualdades mediante sistemas automatizados que, al aprender de datos históricos, codifican y legitiman prácticas discriminatorias con el aura de la objetividad matemática (Noble, 2018; Eubanks, 2018).

Se prevé mayor riesgo de sesgo cuando los algoritmos son entrenados con patrones de datos incompletos; por lo que diversos campos de aplicación se pueden ver afectados en la toma de decisiones que se consideran de mayor implicación en los derechos humanos.

Casos ampliamente citados ilustran el problema: herramientas de evaluación de riesgo penal como COMPAS asociadas a disparidades por raza (Angwin et al., 2016), o sistemas de reconocimiento facial con mayores tasas de error para personas con piel más oscura y mujeres (Buolamwini & Gebru, 2018). Este fenómeno no es accidental: refleja la historia

social contenida en los datos, la subrepresentación de comunidades en el pipeline de datos y desarrollo, y metas de optimización desconectadas de principios de justicia (Crawford, 2021; Raji & Buolamwini, 2019).

Para ello, la ética digital ofrece principios y marcos para orientar el diseño, despliegue y supervisión de sistemas de IA hacia fines socialmente valiosos y respetuosos de los derechos humanos. Compendios internacionales sintetizan principios convergentes: beneficencia, no maleficencia, autonomía, justicia, explicabilidad y responsabilidad (Jobin et al., 2019; Floridi et al., 2020). Iniciativas regulatorias—como la Propuesta de Reglamento Europeo de IA (AI Act)—avanzan hacia un enfoque basado en riesgo, obligaciones de gestión del ciclo de vida, evaluaciones de conformidad, documentación técnica, gobernanza de datos, supervisión humana y transparencia, especialmente para sistemas de alto riesgo (European Commission, 2021).

En paralelo, la *Recomendación de la UNESCO sobre la Ética de la IA* promueve la inclusión, la no discriminación, la diversidad y la protección de derechos como ejes de la gobernanza global (UNESCO, 2021). Estas líneas se refuerzan con herramientas prácticas: auditorías algorítmicas, evaluaciones de impacto en derechos fundamentales, mecanismos de queja y reparación, y controles de desempeño por subgrupos (Raji et al., 2020; Mitchell et al., 2019). La literatura subraya que la ética efectiva requiere infraestructura institucional (políticas, estándares, órganos de supervisión) y capacidad organizacional (equipos interdisciplinarios, formación en ética, procesos de revisión) para pasar de principios a prácticas.

En el contexto latinoamericano, estudios recientes destacan la urgencia de construir marcos regulatorios propios que respondan a las desigualdades estructurales de la región. Luna y Herrera (2024) enfatizan que la gobernanza algorítmica en América Latina debe incorporar principios de soberanía tecnológica y justicia social para evitar la dependencia de modelos regulatorios ajenos a sus realidades socioculturales.

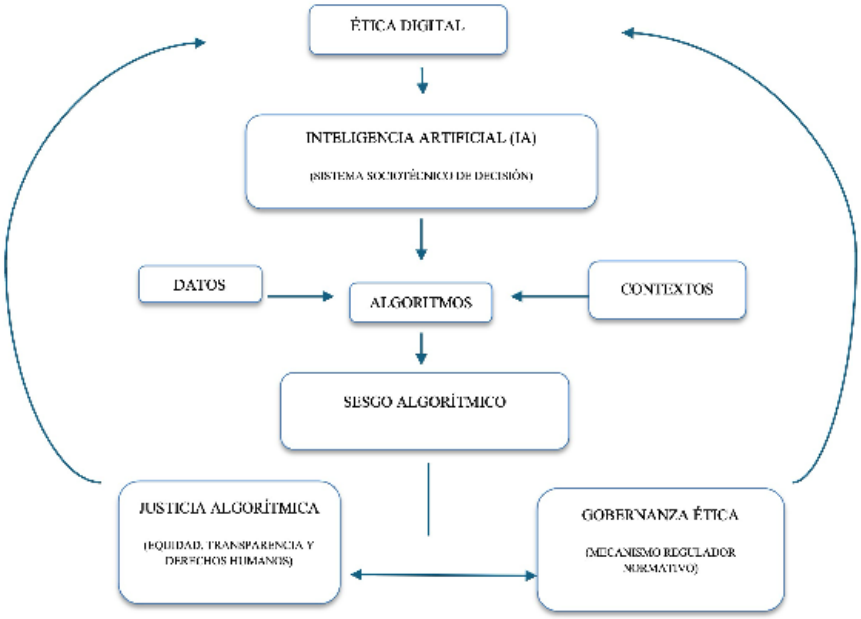
Por tanto, se considera que la IA es sociotécnica: su desempeño y sus efectos dependen tanto de decisiones técnicas como de contextos sociales. El sesgo algorítmico tiene múltiples fuentes y no se resuelve únicamente con más datos; demanda diseño, documentación y evaluación centrados en la equidad y el racismo algorítmico cristaliza desigualdades históricas en sistemas automatizados, exigiendo mitigación técnica y transformación institucional.

La ética digital y la gobernanza de IA requieren principios, normas y herramientas que garanticen justicia, transparencia, rendición de cuentas y derechos humanos. Es decir, se deben establecer las normas y procesos que permitan controlar algoritmos que puedan afectar derechos fundamentales, recursos y toma de decisiones en los diferentes ámbitos, para que se desarrollen de manera justa y responsable.

La Figura 1 presenta el marco relacional de los principales ejes teóricos del estudio. En este modelo, la Ética Digital se ubica como principio rector que orienta el desarrollo y aplicación de la IA, entendida como un sistema sociotécnico. De la interacción entre los datos, los algoritmos y los contextos sociales emergen sesgos algorítmicos que pueden reproducir desigualdades estructurales. Ante estos riesgos, la Gobernanza Ética actúa como un mecanismo regulador mediante normas, auditorías y políticas inclusivas. El resultado de este proceso, encontrándose a la par de la Gobernanza Ética, es la Justicia Algorítmica, entendida como la garantía de equidad, transparencia, responsabilidad y respeto a los derechos humanos en los entornos automatizados.

Figura 1

Marco referencial del surgimiento del sesgo algorítmico



La figura representa la relación conceptual entre la ética digital, la IA, el sesgo algorítmico, la gobernanza ética y la justicia algorítmica, destacando su articulación dentro del marco de la gobernanza ética y la equidad tecnológica.

Resultados

Del diagnóstico global de los sesgos algorítmicos, los resultados de la revisión documental evidencian que la presencia de sesgos algorítmicos constituye una de las problemáticas más complejas y persistentes en el desarrollo contemporáneo de sistemas de IA. En los Estados Unidos, estudios pioneros como los de ProPublica demostraron que el sistema de evaluación judicial COMPAS sobrestimaba el riesgo de reincidencia de personas afrodescendientes y lo subestimaba en individuos blancos (Angwin et al., 2016). Este hallazgo impulsó un debate ético y legal sobre la responsabilidad de los algoritmos en la reproducción de la discriminación institucional, visibilizando la urgencia de mecanismos de rendición de cuentas en la toma de decisiones automatizadas (Barocas & Selbst, 2016; Raji & Buolamwini, 2019).

En el ámbito empresarial, investigaciones como la de Buolamwini y Gebru (2018) mostraron que los sistemas comerciales de reconocimiento facial de IBM, Microsoft y Face presentaban errores de clasificación de género de hasta 35 % en rostros femeninos de piel oscura, en contraste con menos del 1 % en hombres blancos. Estos sesgos se explican, en parte, por la subrepresentación de grupos minoritarios en los conjuntos de datos de entrenamiento, fenómeno que convierte a la IA en un espejo de las desigualdades sociales y raciales existentes (Crawford, 2021; Noble, 2018).

En México, aunque el desarrollo de la IA aún se encuentra en una etapa incipiente, los riesgos asociados a la discriminación algorítmica comienzan a emerger en ámbitos como la seguridad pública, la educación digital y los servicios financieros (Roldán & García, 2023). A diferencia de Estados Unidos, donde se discuten marcos legales específicos como el *Algorithmic Accountability Act* (2022), México carece de una legislación particular sobre la ética o transparencia algorítmica, y se limita a disposiciones generales contenidas en la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, en la Norma Federal de Transparencia y Acceso a la Información Pública, así como, aún en iniciativa, la Ley para la Regulación Ética de la Inteligencia Artificial y la Robótica, y la Ley General de Derechos Humanos Digitales en discusión legislativa (Cámara de Diputados, 2023).

El diagnóstico revela que el país enfrenta tres desafíos principales: ausencia de políticas públicas sobre gobernanza ética de IA; débil infraestructura institucional para auditar algoritmos y evaluar riesgos de discriminación automatizada y escasa formación ética y digital en los profesionales encargados del diseño tecnológico.

Si bien existen iniciativas académicas y gubernamentales que promueven el uso responsable de la IA como la Estrategia Nacional de IA de la Secretaría de Economía (2022), estas carecen de mecanismos de seguimiento, evaluación y sanción (OECD, 2023). Por lo tanto, la distancia entre el discurso y la práctica en materia de ética tecnológica continúa siendo amplia.

A partir del análisis documental comparativo y como resultado del estudio, se identificaron diferencias sustantivas entre México y Estados Unidos en materia de regulación, supervisión institucional y formación ética en inteligencia artificial. La Tabla 2 resume estos contrastes y los vacíos normativos que influyen directamente en la gobernanza algorítmica, evidenciando la necesidad de marcos regulatorios contextualizados a la realidad latinoamericana.

Tabla 2
Comparativo de vacíos normativos y capacidades institucionales en materia de gobernanza algorítmica

Dimensión	Estados Unidos	México	Implicaciones éticas
Marco normativo	Algorithmic Accountability Act (2022); regulaciones sectoriales	Sin ley específica; dependiente de protección de datos personales	Falta de regulación impide mecanismos de auditoría y transparencia
Supervisión institucional	Federal Trade Commission (FTC); AI Ethics Labs	Secretaría de Economía, SECIHTI (sin atribuciones regulatorias)	Escasa capacidad de control ético y técnico
Educación y formación	Programas de ética en ingeniería y ciencias de datos	Escasas asignaturas formales en ética digital	Débil alfabetización ética y tecnológica

Nota. Elaboración propia con base en fuentes de Calo (2021), Roldán y García (2023) y documentos legislativos de ambos países.

Este contraste evidencia que Estados Unidos ha avanzado en la institucionalización de la ética algorítmica, aunque enfrenta críticas por la falta de sanciones efectivas (Calo, 2021). En México, la discusión apenas comienza, centrada en la alfabetización digital y la justicia tecnológica como prioridades para el desarrollo sostenible (Roldán & García, 2023).

Algunas estrategias de mitigación hacia una gobernanza ética e inclusiva que, generadas a partir de los resultados y los hallazgos encontrados, se proponen para mitigar los sesgos algorítmicos y fortalecer la justicia en IA, son:

a) Auditorías éticas y transparencia algorítmica

Las auditorías éticas internas y externas permiten identificar y corregir sesgos antes de la implementación de sistemas de IA. Raji et al. (2020) proponen un modelo de *auditoría continua* que combina revisiones técnicas, participación comunitaria y mecanismos de divulgación pública. Este enfoque refuerza la rendición de cuentas y fomenta la confianza social en las tecnologías.

b) Inclusión y diversidad en el desarrollo tecnológico

La diversidad de género, raza y formación en los equipos de diseño de IA se ha demostrado esencial para reducir sesgos (West et al., 2019). La incorporación de perspectivas interdisciplinares (ética, sociología, derecho) permite cuestionar las decisiones aparentemente *neutrales* del diseño algorítmico y promover sistemas más inclusivos.

c) Alfabetización ética y formación profesional

Promover una educación ética y digital desde la formación universitaria resulta clave para construir una cultura tecnológica responsable (UNESCO, 2021). Los futuros ingenieros y científicos de datos deben adquirir competencias en justicia algorítmica, explicabilidad, y derechos humanos digitales, integradas curricularmente.

A partir de los resultados anteriores, se presenta a continuación una discusión crítica que contrasta los hallazgos con los principales referentes teóricos sobre justicia y ética algorítmica.

Discusión

El análisis comparativo revela que la justicia algorítmica no puede lograrse solo mediante soluciones técnicas, sino que requiere reformas estructurales que aborden las desigualdades sociales preexistentes. En América Latina, la dependencia tecnológica y la asimetría en la producción de datos profundizan los riesgos de replicar modelos de IA diseñados en contextos distintos (Guzmán & Espinosa, 2022).

Los resultados obtenidos coinciden con las reflexiones de Feenberg (2020), quien sostiene que toda tecnología encarna valores sociales y estructuras de poder, por lo que los sesgos algorítmicos deben entenderse como expresiones técnicas de desigualdades estructurales.

En este sentido, la ética de la IA propuesta por Floridi et al. (2020) enfatiza que la justicia algorítmica requiere pasar de los principios a la práctica mediante marcos de gobernanza verificables y mecanismos de rendición de cuentas. Asimismo, Jobin, Ienca y Vayena (2019) advierten que los lineamientos éticos globales solo serán eficaces si se contextualizan cultural y políticamente, lo que resulta particularmente relevante para América Latina, donde la desigualdad digital agrava los riesgos de discriminación automatizada.

Por último, autores como Crawford (2021) y Noble (2018) destacan la necesidad de visibilizar la relación entre datos, poder y exclusión, reforzando el argumento de que la justicia algorítmica no puede limitarse a la corrección técnica del sesgo, sino que debe implicar un replanteamiento ético y político del diseño tecnológico.

La región necesita marcos éticos que reconozcan su diversidad cultural y promuevan la soberanía tecnológica. La adopción de lineamientos como los propuestos por la UNESCO (2021) y la OCDE (2023) representa un punto de partida, pero su eficacia dependerá de la voluntad política y de la participación de la sociedad civil en la supervisión del desarrollo algorítmico.

Este estudio aporta una perspectiva latinoamericana a los debates internacionales sobre justicia algorítmica, al evidenciar que la gobernanza ética de la IA no puede equipararse a modelos europeos o norteamericanos sin considerar las asimetrías estructurales de la región. La propuesta de un modelo basado en auditorías éticas, diversidad en el desarrollo tecnológico y alfabetización ética universitaria representa una contribución original que articula la justicia algorítmica con la agenda de derechos humanos digitales y la inclusión social en América Latina.

Este trabajo presenta algunas limitaciones metodológicas propias de su diseño documental. La revisión se basó principalmente en fuentes en inglés y español, lo que pudo restringir la diversidad lingüística del corpus analizado. Asimismo, la ausencia de datos empíricos impidió examinar directamente las percepciones o experiencias de actores sociales vinculados con el desarrollo de IA en contextos latinoamericanos.

Futuras investigaciones podrían ampliar este análisis mediante estudios empíricos comparativos, entrevistas a diseñadores y legisladores, o análisis de casos locales sobre la aplicación de la justicia algorítmica en el sector público, educativo o judicial. Igualmente,

sería pertinente profundizar en los mecanismos de participación ciudadana en la supervisión tecnológica, así como en la evaluación de políticas públicas de ética digital en América Latina. Estas líneas de investigación permitirán consolidar una agenda interdisciplinaria que articule la teoría con la práctica, fortaleciendo la gobernanza ética de la inteligencia artificial desde una perspectiva humanista y regional.

Conclusiones

El análisis realizado confirmó que la IA no operó como un sistema neutral, sino como un ensamblaje sociotécnico que incorporó y amplificó las desigualdades históricas presentes en los datos y en las instituciones que la desarrollan. Los casos revisados en justicia penal, reconocimiento facial y ámbitos empresariales mostraron que los algoritmos pueden reproducir sesgos raciales, de género y socioeconómicos con apariencia de objetividad, generando riesgos directos para derechos fundamentales como la igualdad de trato, la no discriminación y el debido proceso.

Desde una perspectiva comparativa, el estudio evidenció asimetrías relevantes entre Estados Unidos y México en materia de regulación, capacidades institucionales y formación ética en IA. Mientras que en el contexto estadounidense se han impulsado marcos normativos específicos y mecanismos incipientes de rendición de cuentas, en México la gobernanza algorítmica continúa dispersa en normas generales de protección de datos y transparencia, sin instrumentos robustos para auditar algoritmos ni atribuciones claras de supervisión. Esta brecha refuerza la urgencia de construir marcos regulatorios contextualizados a las realidades latinoamericanas, en los que la justicia social y los derechos humanos digitales ocupen un lugar central.

En el plano teórico, el trabajo aportó un marco relacional que articuló cuatro ejes: ética digital, IA como sistema sociotécnico, sesgo y racismo algorítmico, y gobernanza ética orientada a la justicia algorítmica. Este modelo permitió comprender que los sesgos no se reducen a fallas técnicas corregibles únicamente con más datos, sino que remiten a decisiones de diseño, criterios de optimización y estructuras de poder que se sedimentan en los sistemas automatizados. La propuesta de situar la justicia algorítmica como resultado de la interacción entre ética digital, gobernanza y participación social constituye una contribución conceptual para la discusión regional.

En el plano propositivo, el estudio sugirió un modelo de gobernanza ética sustentado en tres pilares: 1) auditorías éticas y técnicas que permitan evaluar de manera sistemática el desempeño diferenciado de los algoritmos y su impacto en grupos históricamente vulnerados; 2) inclusión y diversidad en los equipos de desarrollo tecnológico, incorporando perfiles de género, raza y disciplinas como derecho, sociología y filosofía de la tecnología; y 3) alfabetización ética y digital como competencia transversal en la formación universitaria, especialmente en ingeniería, ciencias de datos y áreas afines. Estos elementos ofrecen insumos concretos para el diseño de políticas públicas, lineamientos institucionales y programas educativos orientados a una IA más justa, transparente y responsable.

Al mismo tiempo, se reconoció que el diseño documental del estudio implicó limitaciones. El análisis se basó en literatura en inglés y español y no incluyó evidencia empírica directa sobre experiencias de comunidades afectadas por la discriminación algorítmica en América Latina. Estas restricciones abren un campo fértil para futuras investigaciones empíricas que profundicen y tensionen las conclusiones aquí esbozadas.

Entre las principales líneas de investigación futura se identificaron:

1. Estudios empíricos comparativos sobre la implementación de sistemas de IA en sectores como justicia penal, seguridad pública, educación, salud y servicios financieros, con enfoque en cómo se materializan los sesgos y qué mecanismos de resistencia o adaptación emergen desde los actores afectados.

2. Investigaciones jurídico–comparadas que analicen en detalle la recepción, adaptación o resistencia a marcos como el *AI Act europeo* y el *Algorithmic Accountability Act* en contextos latinoamericanos, identificando oportunidades y riesgos de trasplantes normativos acríticos.

3. Diseños participativos y co-creación de mecanismos de auditoría algorítmica con comunidades, organizaciones de la sociedad civil y colectivos vulnerabilizados, a fin de vincular la justicia algorítmica con prácticas de democracia participativa y control ciudadano de la tecnología.

4. Evaluación de políticas y programas de ética digital en universidades, empresas y administraciones públicas, midiendo su impacto real en la reducción de sesgos, la mejora de la transparencia y el fortalecimiento de la rendición de cuentas.

5. Análisis interseccionales que exploren cómo se entrecruzan raza, género, clase, discapacidad y pertenencia indígena en los efectos de los sistemas automatizados, ampliando así la agenda de la justicia algorítmica más allá de los casos paradigmáticos del Norte Global.

Estas proyecciones apuntan a consolidar una agenda de investigación interdisciplinaria y situada que vincule teoría crítica de la tecnología, derecho digital, ciencia de datos y pedagogía crítica. Profundizar en estas líneas permitirá avanzar hacia modelos de gobernanza algorítmica que no solo mitiguen daños, sino que contribuyan a transformar las condiciones estructurales de desigualdad, orientando la inteligencia artificial hacia la equidad, la dignidad y el respeto efectivo de los derechos humanos en América Latina.

Referencias

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*(3), 671–732. https://doi.org/10.15779/Z38BG31

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency* (FAT*), 149–159. https://doi.org/10.1145/3287560.3287576

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research* (PMLR), 81, 1–15. https://proceedings.mlr.press/v81/buolamwini18a.html

Calo, R. (2021). Artificial intelligence policy: A primer and roadmap. *UC Davis Law Review*, *55*(2), 399–434. https://lawreview.law.ucdavis.edu/issues/55/2/articles/55-2_Calo.pdf

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press. https://yalebooks.yale.edu/book/9780300264630/atlas-of-ai/

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214–226. https://doi.org/10.1145/2090236.2090255

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

European Commission. (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence* (Artificial Intelligence Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Feenberg, A. (2020). Technology and the critique of instrumentality. En *Between reason and experience* (pp. 75–94). MIT Press. https://doi.org/10.7551/mitpress/9780262062749.003.0005

Flick, U. (2018). *An introduction to qualitative research* (6th ed.). SAGE Publications.

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics*, *26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Freire, P. (1970). *Pedagogía del oprimido*. Siglo XXI Editores.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. https://doi.org/10.1145/3458723

Guzmán, C., & Espinosa, J. (2022). Ética y justicia algorítmica en América Latina: Retos para la regulación tecnológica. *Revista Latinoamericana de Ciencia, Tecnología y Sociedad*, *17*(2), 85–104. https://doi.org/10.1590/rlcts.v17n2.2022

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *Proceedings of the 8th Innovations in Theoretical Computer Science Conference* (ITCS).https://doi.org/10.1145/3097983.3105811

Luna, D., & Herrera, M. (2024). Ética y regulación algorítmica en América Latina: Desafíos para la gobernanza digital y los derechos humanos. *Revista Latinoamericana de Tecnología y Sociedad*, *19*(1), 88-101.

Llano, F. H. (2024). La ética de los algoritmos.Teoría y derecho: Derecho e inteligencia artificial, No. 37, 74-75. https://doi.org/10.36151/TD.2024.103

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Mitchell, M. (1997). *Machine learning*. McGraw-Hill.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT*), 220–229. https://doi.org/10.1145/3287560.3287596

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5

OECD. (2023). *OECD framework for the classification of AI systems*. Organisation for Economic Co-operation and Development. https://doi.org/10.1787/3c8f24d3-en

Omi, M., & Winant, H. (2014). *Racial formation in the United States* (3rd ed.). Routledge. https://doi.org/10.4324/9780203076804

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. *Crown Publishing Group*. https://doi.org/10.2307/j.ctv1xxs0j

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *AIES '19*, 429–435. https://doi.org/10.1145/3306618.3314244

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Lux, E., Denton, E., & Hanna, A. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (FAT*), 33–44. https://doi.org/10.1145/3351095.3372873

Roldán, M., & García, D. (2023). Gobernanza algorítmica y derechos digitales en México: Desafíos para la ética tecnológica. *Revista Mexicana de Derecho Informático*, *11*(2), 45–63. https://doi.org/10.22201/iij.25940082e.2023.2.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

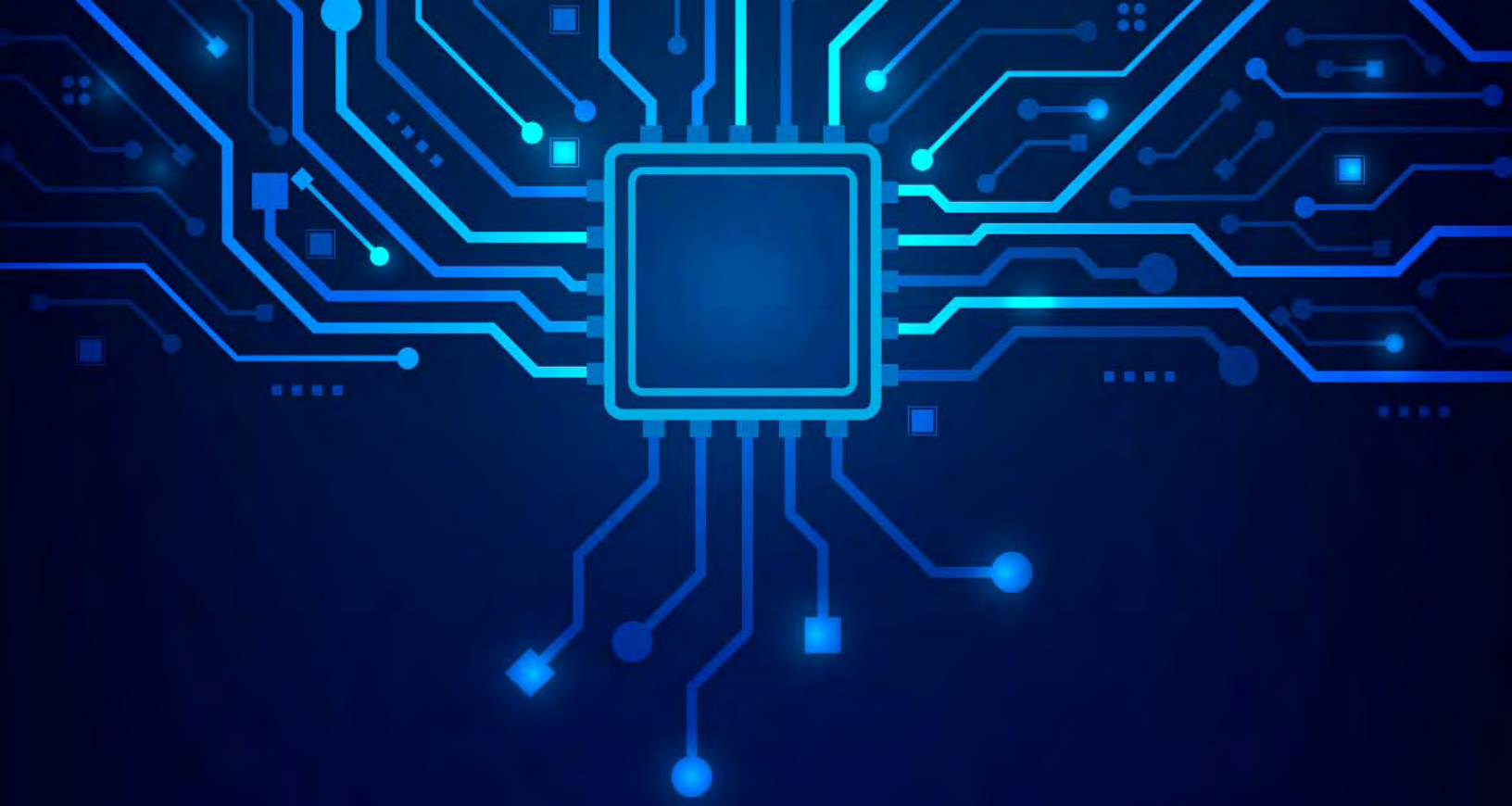
Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE Publications.

Secretaría de Economía. (2022). *Estrategia Nacional de Inteligencia Artificial: México hacia un futuro digital responsable*. Gobierno de México. https://www.gob.mx/se

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). *Fairness and abstraction in sociotechnical systems*. FAT* '19, 59–68. https://doi.org/10.1145/3287560.3287598

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO publishing. https://unesdoc.unesco.org/ark:/48223/pf0000380455

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating systems: Gender, race, and power in AI*. AI Now Institute Report.



05.

Inteligencia artificial generativa y ética profesional en la contaduría pública

Leonardo Eliphas Daza Ramírez

Universidad de Guadalajara, México

ORCID: 0009-0002-3619-8488

Francisco de Jesús Mata Gómez

Universidad de Guadalajara, México

ORCID: 0009-0004-1870-4484

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.05

05.

Inteligencia artificial generativa y ética profesional en la contaduría pública

Introducción

La incorporación de la inteligencia artificial generativa (IAG) en la contaduría pública está transformando de manera importante los procesos de análisis, registro, interpretación y toma de decisiones en la profesión contable. Por tanto, las herramientas que se basan en modelos generativos de lenguaje empiezan a desempeñar un papel relevante tanto en la formación universitaria del licenciado en contaduría pública como en su ejercicio profesional, creando un escenario caracterizado por la digitalización y la automatización cognitiva y la delegación parcial de tareas que antes eran casi al 100% humanas.

Sin embargo, junto con sus beneficios en términos de eficiencia, rapidez y acceso a información, la IAG plantea dilemas éticos de relevancia para la Contaduría Pública. A diferencia de otras innovaciones tecnológicas previas, la IAG no se limita a ejecutar instrucciones programadas, sino que produce recomendaciones, inferencias y soluciones que influyen en decisiones contables, fiscales y financieras con implicaciones legales, económicas y sociales. Este fenómeno problematiza los principios fundamentales de la ética profesional, tales como la integridad, la objetividad, la competencia profesional, la confidencialidad y la responsabilidad social del contador público.

En nexo a lo anterior, este trabajo analiza críticamente el impacto ético de la IAG en la formación y ejercicio profesional del contador público, en lo relativo al juicio profesional y a la toma de decisiones éticas. Asimismo, resulta pertinente cuestionar si los marcos éticos profesionales vigentes, diseñados en tiempos previos al nacimiento de la inteligencia artificial avanzada, son suficientes para enfrentar los dilemas éticos actuales derivados del uso de la IAG. Si bien los códigos de ética siguen siendo un referente normativo indispensable, la complejidad de los entornos digitales exige una reinterpretación crítica de dichos instrumentos, así como su articulación con enfoques contemporáneos de ética de la tecnología, gobernanza algorítmica y responsabilidad profesional socialmente responsable.

Por lo que este estudio tiene como objetivo central analizar, desde un enfoque cualitativo, documental y argumentativo, las implicaciones éticas de la IAG en la formación y el ejercicio profesional del contador público. A partir de la revisión crítica de literatura especializada y de marcos éticos profesionales, se busca identificar los principales dilemas éticos asociados a la IAG, examinar su incidencia en el juicio profesional y valorar la pertinencia de los códigos de ética actuales frente a los avances tecnológicos modernos.

Planteamiento del problema

La contaduría pública es una profesión sustentada en principios éticos: la integridad, la objetividad, la confidencialidad, la competencia profesional y la responsabilidad ético – social, entre otros. Estos principios orientan el ejercicio del contador público y constituyen la base de la confianza social depositada en la información financiera, fiscal, contable y administrativa que produce y valida.

La IAG se ha incorporado en el ámbito contable como una herramienta capaz de automatizar tareas complejas, generar reportes financieros, interpretar normatividad fiscal, apoyar procesos de auditoría y ofrecer recomendaciones basadas en grandes volúmenes de datos e información. Estas aplicaciones cuentan con eficiencia, rapidez y reducción de errores operativos, aunque, también introducen nuevos dilemas éticos que no han sido investigados con profundidad desde una perspectiva de la ética profesional. Entre estos dilemas destacan la posible delegación acrítica del juicio profesional al algoritmo, la opacidad en los procesos de decisión automatizados, los sesgos incorporados en los modelos de IAG, la protección de datos confidenciales y la falta de responsabilidad ética del contador público frente a terceros.

La adopción de la IAG en la contaduría pública plantea interrogantes como: ¿quién es responsable de una decisión contable, fiscal o financiera apoyada en sistemas generativos?, ¿hasta qué punto el contador público cuenta con autonomía ética y moral cuando confía en resultados producidos por algoritmos?, ¿están los marcos o códigos éticos profesionales actualizados para estos nuevos escenarios de la IAG?

Asimismo, en el ámbito de la formación universitaria, la incorporación de la IAG en la enseñanza de la contaduría pública plantea nuevos desafíos. Por ejemplo, existe la problemática del uso indiscriminado de herramientas generativas que no favorecen los procesos formativos, como el pensamiento crítico, el razonamiento ético y la toma de decisiones res-

ponsables, entre otros, si no se integran bajo lineamientos que tengan eticidad profesional. Sin embargo, la literatura especializada muestra poco desarrollo en torno a la ética profesional contable frente a la IAG, particularmente desde enfoques cualitativos y argumentativos que permitan problematizar sus implicaciones.

En este sentido, el problema principal que se analiza hace referencia a la insuficiente reflexión teórica y ética sobre el impacto de la IAG en la formación y ejercicio profesional del contador en México, lo que genera incertidumbre respecto a la preservación del juicio ético, la responsabilidad profesional y la función social del contable en contextos altamente digitalizados. Esta ausencia de análisis ético sistemático limita la construcción de marcos teórico – conceptuales que reorienten tanto la práctica profesional como la formación universitaria, frente a una tecnología que avanza más rápido que las leyes y normas, los códigos de ética y las reflexiones académicas.

Por lo que, se parte de un enfoque cualitativo–argumentativo y se propone contribuir a la discusión teórica sobre la ética profesional en la contaduría pública ante la IAG, analizando oportunidades y desventajas, con el fin de aportar elementos teórico–conceptuales y metodológicos que fortalezcan una adopción crítica, responsable y socialmente aceptada de estas tecnologías en el mundo contable.

Preguntas de investigación

Pregunta principal:

¿Cómo impacta la incorporación de la IAG en el ejercicio de la contaduría pública desde una perspectiva de la ética profesional?

Preguntas específicas:

¿Qué dilemas éticos emergen en la práctica contable y fiscal a partir del uso de herramientas de IAG?

¿En qué medida los códigos de ética profesional vigentes en la contaduría pública resultan suficientes para orientar el uso ético de la IAG?

¿Cuáles son los principales riesgos éticos asociados al uso acrítico de la IAG en procesos contables, fiscales, financieros y de auditoría?

¿Qué implicaciones éticas tiene la adopción de la IAG en la formación universitaria del contador público, particularmente en el desarrollo del pensamiento crítico y ético?

Justificación

La participación de la IAG en materia contable, fiscal y financiera ha superado la capacidad de reflexión teórica y normativa de la ética profesional contable. Si bien existe una creciente producción científica sobre aplicaciones técnicas de la inteligencia artificial en la Contaduría Pública, los estudios que abordan de manera crítica y ética el impacto de la IAG en el juicio profesional, la responsabilidad moral y la función social del contador público continúan siendo limitados en el contexto mexicano.

Desde el punto de vista académico, este artículo intenta contribuir al debate teórico actual sobre ética profesional en la contaduría pública, incorporando un fenómeno emergente que redefine la práctica profesional y al mismo tiempo aportar un enfoque cualitativo, documental y argumentativo, necesario para problematizar dimensiones éticas que no pueden ser reducidas a mediciones empíricas inmediatas.

Además, asocia la ética profesional del contador con marcos conceptuales recientes, ampliando el diálogo interdisciplinario de la contaduría y la ética profesional aplicada y también, intenta fortalecer líneas de investigación emergentes vinculadas con ética de lo fiscal, ciudadanía profesional y responsabilidad social del contador.

La contaduría pública cumple una función estratégica en la transparencia financiera, la justicia fiscal y la confianza pública en las instituciones económicas. La adopción de sistemas de IAG en procesos contables y fiscales puede generar beneficios significativos en eficiencia y precisión; Sin embargo, ello también conlleva riesgos éticos y morales que afectan a contribuyentes, empresas, gobiernos y a la sociedad en general.

La justificación social de este estudio, se encuentra en el hecho de que, se analizan los riesgos de deshumanización y desplazamiento de la responsabilidad ética en decisiones contables apoyadas en algoritmos y se advierte sobre el peligro de una dependencia tecnológica acrítica, que puede debilitar la rendición de cuentas y la ética profesional, y, de paso contribuye a la formación de una cultura ética digital en la profesión contable, necesaria para enfrentar los retos de la economía digital y la automatización inteligente.

Objetivo general

Analizar, desde un enfoque cualitativo, documental y argumentativo, las implicaciones éticas de la Inteligencia Artificial Generativa en la formación y ejercicio profesional del contador público.

Objetivos específicos

Identificar los principales dilemas éticos que surgen con la incorporación de la Inteligencia Artificial Generativa en la práctica de la contaduría pública.

Analizar cómo el uso de la IAG incide en el ejercicio del juicio profesional y en la toma de decisiones éticas del contador público.

Examinar críticamente la pertinencia y suficiencia de los marcos éticos profesionales actuales frente a los desafíos tecnológicos que plantea la IAG.

Supuestos teóricos

La ética profesional constituye un eje insustituible del ejercicio de la contaduría pública, por lo que ninguna tecnología puede sustituir el juicio moral y la responsabilidad social del contador público.

La IAG no es éticamente neutral, ya que incorpora decisiones humanas, sesgos algorítmicos y criterios implícitos que influyen en los resultados contables, fiscales y financieros que produce.

El uso acrítico de la IAG tiende a debilitar el juicio profesional del contador público, favoreciendo una dependencia tecnológica que puede erosionar la autonomía ética y la responsabilidad del contador.

Los códigos de ética profesional vigentes resultan necesarios pero insuficientes para enfrentar los dilemas éticos emergentes derivados del uso de la IAG.

Marco teórico

La ética profesional en la contaduría pública se encuentra fundamentada en principios reconocidos como son: la integridad, la objetividad, la competencia profesional, la confi-

dencialidad y el comportamiento profesional. Estos principios han sido sistematizados y difundidos por organismos internacionales como la Federación Internacional de Contadores (IFAC, 2023), que concibe la ética no solo como un conjunto de normas, sino como un compromiso social del contador público y el interés de la sociedad.

En México existe el Código de Ética Profesional del (Instituto Mexicano de Contadores Públicos [IMCP], 2024), el cual se encuentra armonizado con el IFAC y hace referencia a la importancia de las salvaguardas, la confidencialidad, y la objetividad e independencia, así como el fortalecimiento de la competencia profesional continua, principios que resultan aplicables al uso de IAG (IMCP, 2024).

Par tener un acercamiento teórico-conceptual de la pertinencia de la IAG en relación con la ética profesional del contador público en México, se construyó un breve resumen del Código de Ética Profesional del IMCP (2024), el cual, constituye el marco normativo fundamental que orienta la conducta ética del profesional de la contaduría pública y tiene como propósito central salvaguardar el interés público, garantizar la confianza social en la profesión contable y establecer principios que regulen el ejercicio profesional.

La estructura del Código de Ética del IMCP tiene cinco principios fundamentales, un marco conceptual de amenazas y salvaguardas, y disposiciones específicas aplicables a distintos ámbitos del ejercicio profesional, los principios son:

a) Integridad: el contador público debe actuar con honestidad y rectitud en todas sus relaciones profesionales y de negocios. En el contexto de la IAG, este principio implica que el profesional no debe presentar como propios ni infalibles los resultados generados por sistemas inteligentes, ni ocultar el uso de dichas herramientas cuando ello sea relevante para la toma de decisiones.

b) Objetividad: exige que el contador no permita que prejuicios, conflictos de interés o influencias indebidas afecten su juicio profesional. Ante la IAG, este principio tiene especial importancia, en razón de que los algoritmos pueden incorporar sesgos implícitos; por ello, es recomendable que el contador cuente con una actitud ética para evaluar críticamente los resultados automatizados y no aceptarlos de manera acrítica.

c) Competencia profesional y debido cuidado: obliga al contador a mantener conocimientos y habilidades actualizados, así como a actuar con diligencia. En el ámbito de la IAG, este principio se amplía hacia la comprensión básica del funcionamiento,

alcances y limitaciones de las herramientas tecnológicas inteligentes, de modo que su uso sea éticamente responsable.

d) Confidencialidad: el contador debe respetar la confidencialidad de la información obtenida como resultado de relaciones profesionales. La utilización de la IAG plantea riesgos adicionales en materia de protección de datos, por lo que este principio exige especial cuidado en el manejo de información financiera, fiscal y personal.

e) Comportamiento profesional: implica cumplir con las leyes y reglamentos aplicables y evitar cualquier conducta que desacredite a la profesión. En relación con la IAG, este principio demanda que el contador no delegue de manera inética decisiones profesionales en sistemas tecnológicos ni actúe de manera que se diluya su responsabilidad legal frente a clientes, autoridades o la sociedad.

Marco conceptual: amenazas y salvaguardas

El Código del IMCP (2024), incorpora un marco conceptual que obliga al contador público a: identificar amenazas al cumplimiento de los principios éticos, evaluar la significatividad de dichas amenazas, y aplicar salvaguardas adecuadas para eliminarlas o reducirlas a un nivel aceptable.

Desde la perspectiva de la IAG, pueden identificarse nuevas amenazas éticas emergentes, por ejemplo: Dependencia excesiva de sistemas automatizados, reducción del juicio profesional humano y opacidad en los procesos de decisión algorítmica, riesgos en la confidencialidad y uso indebido de información desplazamiento implícito de la responsabilidad ética hacia la tecnología.

Alineación del Código de Ética del IMCP con la IAG

Aunque el Código de Ética del IMCP no fue diseñado para regular la IAG, su estructura y principios permiten una interpretación ética aplicable a estos nuevos contextos tecnológicos. El Código deja implícito que:

a) La responsabilidad ética es siempre del contador público, independientemente de las herramientas tecnológicas que utilice.

b) El juicio profesional no puede ser sustituido por sistemas automatizados.

c) La tecnología debe estar subordinada al interés público y a la ética profesional.

d) El uso de herramientas innovadoras exige mayor, no menor, responsabilidad moral.

Por lo que se deduce que, el Código de Ética del IMCP ofrece una base fundamental para orientar el uso ético de la IAG, pero al mismo tiempo evidencia la necesidad de reinterpretaciones contemporáneas, guías complementarias y reflexiones académicas que ayuden a actualizar su aplicación.

En nexo a lo anteriormente mencionado, la Tabla 1 presenta una comparativa entre el Código de Ética Profesional del Contador Público (2024), la Federación Internacional de Contadores – IFAC–Código Internacional de Ética para Contadores Profesionales Internacionales ([IESBA] 2024) y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2022).

Tabla 1

Comparativa entre el Código de Ética Profesional del Contador Público (2024), la IFAC–IESBA (2024) y la UNESCO e IA (2022)

	IMCP–México	IFAC–IESBA	UNESCO
Dimensión	Código de Ética Profesional que establece normas éticas mínimas para la práctica contable en México y el interés público	Código Internacional de Ética que establece el estándar ético global para contadores, centrado en responsabilidad de interés público	Recomendaciones sobre la Ética de la Inteligencia Artificial (IA)
Propósito principal	La práctica contable en México y el interés público	Establecer el estándar global para contadores, centrado en responsabilidad de interés público	Definir el estándar global de la ética de la IA basado en derechos humanos
Principios clave	Integridad, objetividad, competencia y debido cuidado confidencial al comportamiento profesional	Los mismos cinco principios fundamentales	Principios como proporcionalidad/no daño, equidad/no discriminación, privacidad, transparencia/ explicaΩΩbilidad, responsabilidad, supervisión humana, sostenibilidad, alfabetización
Enfoque metodológico	Marco conceptual: identificar-evaluar-responder a amenazas; aplicar salvaguardas	Marco conceptual: equivalente (amenazas y salvaguardas) y estructura por partes/ secciones	Enfoque de ciclo de vida de la IA (diseño despliegue, uso y aplicación), Gobernanza adaptativa

Tabla 1

Comparativa entre el Código de Ética Profesional del Contador Público (2024), la IFAC–IESBA (2024) y la UNESCO e IA (2022)

	IMCP–México	IFAC–IESBA	UNESCO
Tipo de amenazas/ riesgos	Amenazas de cumplimiento de principios (Ejemplo: interés personal, intimidación…)	Categorías de amenazas (interés propio, autorrevisión abogacía, familiaridad, intimidación) y salvaguardas.	Riesgos éticos sistémicos; opacidad, daño, vigilancia, afectación de derechos, impactos ambientales; exige control y responsabilidad
Transparencia / explicabilidad	Implícita en integridad, debido cuidado y comportamiento profesional.	Similar: integridad y debido cuidado; gestión ética del juicio profesional	Explícita: Transparencia y explicabilidad como condición para derechos y responsabilidad
Confidencialidad de datos	Confidencialidad como principio. (protección de información profesional)	Confidencialidad como principio fundamental.	Derecho a la privacidad y protección de datos como principio rector en IA.
Competencia profesional	Actualización permanente y debido cuidado (aplicable a herramientas digitales/IA)	Competencias y debido cuidado el cual exige capacidades para evaluar trabajos y riesgos.	Alfabetización/conciencia en IA y capacidades institucionales para uso responsable.

Nota. Elaboración propia con datos e información del Código de Ética del IMCP (2024); el Código Internacional para Contadores Profesionales (IFAC – IESBA, (2024)[1] y UNESCO (2022) en la obra Ética de la Inteligencia Artificial, publicada por el Observatorio Global de Ética y Gobernanza de la IA.[2]

Por otra parte, el corpus analítico se construyó con artículos científicos arbitrados e indexados para conformar la revisión teórico – documental, identificando fuentes relevantes, principalmente de los últimos cinco años. La Tabla 2 presentan los hallazgos como una matriz.

1 https://www.ethicsboard.org/iesba-code

2 https://www.unesco.org/en/artificial-intelligence/recommendation-ethics El Observatorio Global de Ética y Gobernanza de la IA tiene como objetivo: proporcionar un recurso global para que los responsables de políticas, los reguladores, los académicos, el sector privado y la sociedad civil encuentren soluciones a los desafíos más urgentes que plantea la Inteligencia Artificial.

Tabla 2

Matriz de análisis del corpus documental sobre ética, contaduría pública e IAG

Autor y año	Objeto de análisis	Enfoque / tipo de estudio	Principales aportes teóricos	Vinculación con ética profesional e IAG
Ballantine, Boyce & Stoner (2024)	IA en la educación contable	Revisión crítica	Identifica oportunidades didácticas y pedagógicas, además, de riesgos éticos de la IA en la formación del contador público	Evidencia tensiones entre automatización, juicio profesional y ética del docente
Barraés-Herrera (2022)	Metaversos y educación virtual	Teórico–reflexivo	Analiza entornos virtuales inmersivos y sus implicaciones educativas	Aporta contexto tecnológico para dilemas éticos futuros en contaduría digital
Becerra Peña (2023)	Digitalización y tributación (BEPS)	Reflexión jurídico–fiscal	Expone desafíos éticos y normativos de la economía digital	Refuerza la necesidad de la ética profesional ante las tecnologías emergentes
Bustamante Morán et al. (2025)	Percepción docente sobre IA	Estudio empírico (perceptual)	Identifica actitudes, miedos y expectativas docentes frente a la IA	Relaciona ética profesional con adopción crítica de la IA en la educación y formación del contador
Bolaño-García & Duarte-Acosta (2024)	IA en la educación	Revisión sistemática	Sistematiza beneficios y riesgos del uso educativo en la IA	Sustenta la necesidad de marcos éticos transversales
Camargo Arboleda & Herrara Rocha (2025)	IA y gamificación en educación contable	Mixto / aplicado	Evidencia mejoras pedagógica con los riesgos de dependencia tecnológica	Relevante para discutir ética y formación del juicio profesional
Chicaiza et al. (2022)	Realidad aumentada y aprendizaje	Aplicativo–educativo	Analiza tecnologías emergentes en procesos de interaprendizaje	Antecede a dilemas éticos que hoy se profundizan con la IA

Tabla 2

Matriz de análisis del corpus documental sobre ética, contaduría pública e IAG

Autor y año	Objeto de análisis	Enfoque / tipo de estudio	Principales aportes teóricos	Vinculación con ética profesional e IAG
Daza Mercado et al. (2006)	Ética profesional en la Contaduría Pública	Teórico–formativo	Fundamenta la ética como el eje del ejercicio profesional de la contaduría pública	Base conceptual para articular ética profesional e IAG
Daza-Ramírez (2025)	Privacidad y plataformas digitales	Analítico-conceptual	Propone marcos de integridad y cumplimiento en entornos digitales	Relevante para confidencialidad y uso ético de la IAG
González, S. (2023)	Metaverso y riesgos en educación superior contable	Teórico-aplicado	Identifica riesgos y dilemas éticos contables y antifraude en entornos virtuales	Vincula la ética profesional con tecnologías inmersivas
Palermo (2024)	IA en contabilidad empresarial	Divulgación académica	Describe impactos operativos y estratégicos de la IAG	Reafirma la necesidad la ética profesional en la adopción de herramientas tecnológicas
Rodríguez García et al. (2024)	Percepción estudiantil sobre IA	Estudio empírico	Evidencia adaptación y riesgos percibidos por estudiantes	Vincula ética, formación y uso responsable de la IAG
Villalba Quinta (2024)	IA y profesión contable	Reflexión teórica	Analiza innovación y transformación profesional	Apoya discusión ética sobre redefinición del rol contable
Wanden-Berghe (2023)	Blockchain e IA contable	Investigación académica	Propone disrupción de la partida doble hacia sistemas inteligentes	Plantea retos y dilemas éticos estructurales en la información contable

Nota. Elaboración propia con datos e información de la revisión de literatura sobre ética, contaduría e IAG. En esta tabla se presenta una sistematización rigurosa y no simple listado de fuentes, y se muestra cómo el debate sobre IA, metaverso y educación converge en la ética profesional del contador público.

Una de las principales razones de construir este corpus documental y analizarlo, fue para crear tres grandes líneas para el estudio: 1) IAG en educación contable; 2) ecosistemas digitales inmersivos (metaverso/realidad aumentada), y 3) transformación digital de la contaduría y del entorno fiscal (incluyendo economía digital, BEPS, *blockchain* e IA)[3].

En cuanto a la primera línea: la IAG en la educación contable y en la revisión crítica la plantea, Ballantine, Boyce y Stoner (2024) que aportan una discusión relevante: la IAG genera dilemas entre la formación técnica y la formación ética, al desplazar capacidades del juicio profesional. Bustamante Morán et al. (2025) y Rodríguez García et al. (2024) hacen referencia a trabajos de percepción y plantean un panorama en el que académicos y estudiantes suelen adoptar posturas ambivalentes: valoran los beneficios, pero temen por la calidad del aprendizaje, la autoría, y la confiabilidad de resultados.

La segunda línea refiere al metaverso, la realidad aumentada y digitalización inmersiva, en el que los estudios sobre metaverso y realidad aumentada de (Barráez-Herrera, 2022; González, 2022; López-Belmonte et al., 2023; Olivarría et al., 2022; Chicaiza et al., 2022); describen oportunidades educativas, pero también abordan el tema de la ética como "implicación" general, más que como disciplina aplicada a decisiones contables/fiscales. Otra perspectiva la señala González (2023), al vincular al metaverso con prospectiva complejas en materia contable y de lineamientos antifraude, mostrando que lo inmersivo puede aumentar opacidad, nuevas modalidades de fraude y problemas de evidencia. No obstante, estos trabajos se enfocan más en el "entorno digital" que en la IAG como tecnología cognitiva de producción de contenidos/decisiones, pero, nos ayudan a explicar por qué la ética contable enfrentará circunstancia de mayor complejidad informacional.

La tercera línea explicita la transformación profesional contable y la fiscalidad de la economía digital, la cual es estudiada por autores como Becerra Peña (2023), Villalba Quinta (2024), Palermo (2024), Wanden-Berghe (2023). En síntesis, sostienen que la digitalización modifica prácticas de (trazabilidad, seguridad, evidencia, auditoría continua, nuevos activos), y que el profesional contable deberá reconfigurar competencias. Sin embargo, la pregunta central sigue siendo ética: ¿cómo se preserva la responsabilidad profesional cuando las decisiones se coproducen con sistemas inteligentes? El trabajo de Daza-Ramírez (2025)

3 La matriz permitió: clasificar las fuentes según su enfoque teórico–conceptual, vincular cada documento con las categorías analíticas previamente definidas, identificar aportes teóricos, dilemas éticos y vacíos normativos y relacionar los hallazgos documentales con los supuestos teóricos del estudio. De este modo, la matriz no operó únicamente como un recurso descriptivo, sino como una herramienta interpretativa que facilitó el análisis crítico y la construcción argumentativa.

sobre privacidad en plataformas digitales aporta un puente importante para argumentar la importancia de la integridad y la eticidad.

En nexo a lo anterior, este *corpus* teórico–conceptual aporta una serie de reflexiones sobre las percepciones y riesgos contables emergentes, pero ofrece poco sobre cómo traducirlo en salvaguardas concretas para el contador público ante la IAG[4]. Por eso es estratégico incorporar el Código de Ética del IMCP (2024) y el marco formativo nacional (Daza Mercado et al., 2006) como andamiaje normativo para operacionalizar lineamientos.

Categorías analíticas del corpus teórico documental (2022 – 2026)

Las categorías analíticas[5] construidas que se presentan a continuación, nacen de la matriz de análisis del corpus documental sobre ética, contaduría pública e IAG (tabla 2) y fueron definidas de manera teórica y deductiva, a partir del marco teórico–conceptual del estudio y de los supuestos teóricos planteados. Entre ellas se incluyeron, entre otras: Ética profesional del contador público en contextos tecnológicos; IAG como tecnología no neutral; Juicio profesional y autonomía y dilemas éticos emergentes en la práctica contable.

Categoría 1. Ética profesional del contador público en contextos de la IAG

Definición analítica: conjunto de principios, valores, deberes y responsabilidades que orientan el ejercicio profesional del contador público, en particular, frente a la automatización, digitalización y la IAG.[6]

4 Falta una buena conexión entre la ética profesional (principios/amenazas/salvaguardas) y casos de uso de IAG en contaduría mexicana.

5 Las categorías analíticas no son variables (como en estudios cuantitativos), sino ejes interpretativos y funcionan como andamiajes conceptuales para: organizar el corpus documental, interpretar críticamente los textos, articular los supuestos teóricos con los hallazgos argumentativos.

6 Se entiende a la ética como eje estructural del ejercicio profesional, considerando la responsabilidad social, interés público y confianza y refuerza el supuesto de que ninguna tecnología sustituye el juicio moral

Categoría 2. Inteligencia Artificial Generativa como tecnología no neutral

Definición analítica: la IAG se entiende como un sistema sociotécnico que incorpora decisiones humanas, supuestos epistemológicos, sesgos algorítmicos y criterios normativos implícitos.[7]

Categoría 3. Juicio profesional y autonomía ética del contador público

Definición analítica: capacidad del contador público para deliberar, evaluar contextos, interpretar normas y asumir decisiones responsables, incluso cuando utiliza herramientas avanzadas de IAG.[8]

Categoría 4. Dilemas éticos emergentes en la práctica contable con IAG

Definición analítica: situaciones problemáticas en las que el uso de la IAG genera conflictos entre valores profesionales, normas técnicas, expectativas institucionales y responsabilidades sociales.[9]

Categoría 5. Marcos éticos profesionales frente a la IAG

Definición analítica: conjunto de códigos, normas y lineamientos éticos profesionales actuales que permiten regular la conducta del contador público ante las nuevas herramientas tecnológicas, como es el caso de la IAG.[10]

Categoría 6. Formación ética y pensamiento crítico en la educación contable

Definición analítica: procesos formativos orientados al desarrollo del pensamiento crítico, la deliberación de la ética aplicada y metaética, además de la comprensión de riesgos y límites de la IAG en la formación universitaria del contador público.[11]

---

7 La IAG no es solo herramienta, sino actor mediador ya que, realza una producción automatizada de resultados contables, fiscales y financieros, aunque también puede aportar riesgos de opacidad, sesgo y delegación acrítica.

8 Hace referencia al juicio profesional vs. automatización, al riesgo de dependencia tecnológica y al dilema entre eficiencia técnica y reflexión ética.

9 Refiere a la autoría y responsabilidad de resultados generados por IA, la confidencialidad en el uso de datos y a la transparencia, explicabilidad y rendición de cuentas.

10 Esta categoría es importante porque reflexiona sobre el desfase entre regulación ética y desarrollo tecnológico y nos invita a pensar sobre el que los códigos vigentes son necesarios pero insuficientes.

11 Esta categoría nos recuerda la necesidad de realizar una reconceptualización curricular y plantear una dimensión prospectiva y educativa, en la enseñanza y práctica de la ética profesional del licenciado en contaduría pública.

En síntesis, se afirma con fundamento en los documentos analizados y las categorías analíticas construidas, que la ética profesional continúa siendo el eje estructural del ejercicio de la Contaduría Pública, incluso en contextos de alta automatización e IA, en virtud de la confianza pública, la responsabilidad social y el interés general no pueden ser delegados a sistemas algorítmicos, por avanzados que estos sean.

En este sentido y desde la perspectiva de la IAG, esta se entiende como una herramienta de apoyo y no como un sustituto del juicio moral del profesional. Este hallazgo confirma el supuesto teórico de que ninguna tecnología puede reemplazar la deliberación ética ni la responsabilidad del contador público, sino que, por el contrario, exige una mayor conciencia ética frente a los riesgos de automatización acrítica. La matriz evidenció que esta condición sociotécnica de la IAG plantea dilemas éticos relacionados con la transparencia, la explicabilidad y la atribución de responsabilidades.[12]

Otro de los hallazgos relevantes del corpus es la advertencia sobre el riesgo de debilitamiento del juicio profesional derivado del uso acrítico de la IAG. Este resultado es consistente con el supuesto teórico que plantea que la confianza excesiva en sistemas automatizados puede erosionar la autonomía profesional y generar procesos de "desresponsabilización algorítmica", donde el contador pudiese atribuir a la tecnología decisiones que le competen a la ética y a las normas en materia legal.

Cabe mencionar que, los resultados derivados de la matriz de análisis confirman los supuestos teóricos del estudio y permiten sostener que la incorporación de la IAG en la Contaduría Pública no constituye un desafío técnico, sino un problema ético de fondo.

Y, continuando con el marco teórico en relación con la ética profesional del licenciado en Contaduría Pública, ésta se concibe como una práctica reflexiva, en la que el juicio moral del profesional resulta muy importante para interpretar y aplicar normas, evaluar contextos y asumir responsabilidades. Desde esta perspectiva, el contador no es un ejecutor técnico, sino un agente ético cuyas decisiones tienen efectos económicos, sociales y fiscales de amplio alcance (Daza Mercado, 2022).

Por tanto, la IAG en este trabajo se entiende como un conjunto de sistemas capaces de producir textos, análisis, predicciones y recomendaciones a partir de grandes volú-

12 La discusión se puede desplazar a la pregunta sobre si la IAG es "buena" o "mala", hacia la reflexión sobre quién decide, con qué criterios y bajo qué marcos éticos se implementa, reforzando el supuesto de que la tecnología no es éticamente neutral.

menes de datos y modelos de aprendizaje profundo. A diferencia de otras tecnologías de automatización, la IAG participa de manera directa en procesos cognitivos tradicionalmente humanos, como el razonamiento, la interpretación normativa y la generación de juicios.[13]

Diversos autores señalan que la IAG no es éticamente neutral, ya que incorpora supuestos, valores y sesgos derivados de sus datos de entrenamiento y de las decisiones de diseño de sus desarrolladores (Floridi et al., 2021; UNESCO, 2023). En el ámbito contable, esto implica que los resultados generados por IAG deberán ser evaluados críticamente y no asumidos como objetivamente correctos o moralmente neutros.

Por otra parte, la ética de la IAG ha evolucionado en los últimos años hacia enfoques centrados en la responsabilidad, la transparencia y la rendición de cuentas. Organismos como la Organización para el Desarrollo Económico y la Cooperación (OCDE, 2021 y la UNESCO, 2023) han subrayado que la responsabilidad última de las decisiones apoyadas en IA recae en los seres humanos[14] y no en los sistemas tecnológicos.

Por este motivo, entre otros, la ética profesional debe integrarse como un eje transversal en la formación del contador público, incorporando la reflexión sobre el uso responsable de tecnologías inteligentes, la toma de decisiones éticas en entornos digitales y la preservación del juicio profesional humano frente a la automatización, construyendo espacios donde se use la reflexión ética, con ayuda de la IAG.

Además, es recomendable crear nuevas asignaturas que permitan de manera transversal (es decir, que contengan al interior de sus unidades y objetivos educacionales, el tema de la eticidad y no como una materia aislada y opcional, como sucede actualmente), Dichas materias o asignaturas pudiesen ser: contabilidad digital y uso de *software* ERP y automatización contable, fiscalidad y tributación digital,[15] fiscalidad cuántica, auditoría digital

13 La IAG no es una amenaza para la contaduría pública, sino un catalizador de cambio. El verdadero reto se encuentra en formar contadores con una perspectiva crítica, ética y adaptativa, que sepan usar estas tecnologías como herramientas de apoyo, sin perder de vista el sentido humano, social y ético de la profesión.

14 La Suprema Corte de Justicia de la Nación, determinó que las obras generadas exclusivamente mediante IA (sin intervención creativa humana significativa) no están protegidas por derechos de autor en México. Esto porque la Ley Federal del Derecho de Autor reconoce como autores solo a personas físicas, y requiere originalidad ligada al intelecto humano. https://www.ecija.com/actualidad-insights/discusion-de-la-scjn-sobre-el-caracter-de-dominio-publico-de-las-obras-generadas-con-ia/?utm_source=chatgpt.com

15 El hecho imponible en el metaverso, el IVA e ISR en transacciones inmersivas, plataformas de control fiscal aplicadas al metaverso y mecanismos de cooperación internacional para evitar doble tributación en transacciones virtuales, entre otros temas y subtemas.

y forense,[16] forensia digital,[17] auditoría inmersiva, y, agregar asignaturas relacionadas con: ética profesional y algoritmos,[18] ética fiscal y responsabilidad social digital, metodología de la Investigación con IAG, entre otras. (Comunicación personal con Chat GPT 12 de enero 2026).

Inteligencia artificial generativa y la ética profesional del contador público

En la enseñanza de la profesión contable desde la perspectiva de la eticidad para el ejercicio profesional y reflexionando en prospectiva, habrá que tomar en cuenta escenarios actuales y futuros, que tiene relación con la economía digital y el metaverso que plantean problemáticas inéditas.

Los estudiantes que hoy se preparan, en diez años estarán ante escenarios donde se realizarán auditorías virtuales en entornos inmersivos; existirán transacciones contables basadas en activos digitales, criptomonedas y contratos inteligentes; la enseñanza y el ejercicio profesional se presentará en entornos de plataformas de realidad aumentada y realidad virtual.

En nexo a lo anterior, el mercado ocupacional del contador público para los próximos años se amplía de manera extraordinaria, en relación con la economía virtual y el metaverso, por motivos de las dinámicas de las transacciones económicas, fiscales, financieras y contables dentro de entornos virtuales y, además, considerando aspectos éticos, jurídicos, normativos y tecnológicos en los que deberá participar.

16 Refiere a aplicar técnicas de auditoría digital en la detección, prevención e investigación de fraudes y delitos financieros, mediante herramientas forenses tecnológicas, considerando el Marco legal nacional e internacional (Código Fiscal, Ley Antilavado, ISO 27037), las tipologías de fraudes electrónicos y ciber-delitos financieros, utilizando Software de análisis forense (EnCase, FTK, Autopsy, herramientas de blockchain analytics), métodos de trazabilidad en operaciones financieras electrónicas y aplicada también a delitos fiscales en la economía digital, entre otros.

17 Procedimientos estandarizados de auditoría en entornos inmersivos, Modelos de evidencia digital para peritajes en juicios fiscales mexicanos, Desarrollo de herramientas de auditoría continua en tiempo real aplicadas al metaverso. Análisis de ciberseguridad financiera con casos empíricos en México.

18 Reflexionar sobre los dilemas éticos que enfrentan los contadores públicos en un entorno dominado por algoritmos e inteligencia artificial, desarrollando criterios deontológicos para el ejercicio profesional.

Por lo que, el contador del futuro será un profesional híbrido con sólidas bases tradicionales de contabilidad,[19] finanzas,[20] auditoría[21] y fiscalidad,[22] con competencias en IA, blockchain, metaverso y ciberseguridad y, con un fuerte compromiso en ética y responsabilidad social digital, capaz de desenvolverse en realidades inmersivas y entornos fiscales globales. (Chat GPT, comunicación personal 12 de enero de 2026).

Se observa una tendencia creciente en México, encaminada hacia la formación en alfabetización algorítmica y ética digital en el mundo de la Contaduría Pública, por lo que, es indispensable que se incorporen competencias en uso responsable de IAG, análisis de sesgos y validación de resultados, fortaleciendo el juicio profesional como rasgo distintivo del contador.

La evidencia documental sugiere que la IAG constituye una evolución híbrida en la educación contable, impulsando la transición desde modelos tradicionales hacia ecosistemas digitales donde la creatividad, la ética, la gamificación y la interpretación crítica, adoptan mayor relevancia que la simple memorización de las normas.

Y, como reflexión futurista, es posible imaginar que dentro de 50 años, las Universidades no “impartirán” licenciaturas de “Contaduría Pública”, sino de la “Ciencia Ética del sistema financiero–fiscal y de la gobernanza algorítmica”, y los profesionales usarán interfaces neuronales y modelos generativos de interpretación económica.[23] Pero el eje vertebral de

19 El contador deberá tener competencias en realidad aumentada, realidad virtual y metaverso aplicados a la contaduría; Alfabetización avanzada en IA generativa, big data y analítica predictiva, entre otros.

20 Será gestor de información financiera en entornos híbridos y manejo de contabilidad electrónica avanzada, blockchain y smart contracts, con especialización en criptoactivos, NFTs y bienes virtuales como parte de los estados financieros, etcétera.

21 El contador del futuro realizará ejercicios de auditorías digitales en tiempo real, utilizando herramientas de inteligencia artificial, podrá realizar auditoría inmersiva dentro de plataformas de metaverso y tendrá especialización en ciberseguridad financiera, prevención de fraudes digitales y trazabilidad en blockchain.

22 El LCP tendrá dominio de la fiscalidad electrónica y cumplimiento en plataformas digitales, conocimiento de tratados internacionales de tributación digital (OCDE, G20, BEPS 2.0); Capacidad de asesorar en estrategias fiscales de empresas inmersas en el metaverso y la economía colaborativa global.

23 Los asientos contables serán generados por algoritmos con verificación criptográfica y los estados financieros serán dinámicos, interactivos y predictivos. Los auditores humanos serán reemplazados por auditores algorítmicos y de trazabilidad digital, capaces de rastrear miles de millones de transacciones por segundo, cuidando respetar los principios constitucionales de proporcionalidad, equidad y transparencia. Vale la pena plantearse en este contexto del futuro, algunos dilemas éticos: ¿Puede un algoritmo firmar un dictamen financiero? ¿Quién es responsable cuando una inteligencia generativa comete un error contable o fiscal? ¿Cómo se garantiza la independencia del juicio profesional frente a sistemas autónomos?

la contaduría, seguirá siendo la misma: dar fe, con ética y rigor, de la verdad económica y social, de lo técnicamente correcto y lo éticamente legítimo.[24]

Método de investigación

El estudio se desarrolla con un enfoque metodológico documental cualitativo–argumentativo y realiza un análisis crítico e interpretativo sobre la ética profesional en la Contaduría Pública ante la IAG, en donde se examinan diferentes marcos conceptuales y se logra sistematizar fuentes secundarias verificables publicadas principalmente entre el periodo de 2022 a 2026.

Este enfoque resulta pertinente en virtud de que el objeto de estudio no se centra en la medición de variables cuantificables, sino en la interpretación, problematización y construcción de un corpus teórico relacionado con los ámbitos contable, fiscal y ético en el contexto de la IA.

El corpus estuvo conformado por artículos científicos indexados y arbitrados de diferentes fuentes reconocidas, como: *Dialnet*, *Google Scholar*, *SciELO*, *Redalyc*, *Latindex* y *ResearchGate;* y publicaciones como libros de texto, códigos y publicaciones internacionales de OCDE, UNESCO, IFAC y el IMCP. Excluyendo artículos de divulgación sin sustento metodológico o sin referencias completas, ajenas al tema objeto de estudio de este artículo.

La selección de las fuentes se realizó con base a los siguientes criterios: pertinencia temática, vinculada con la IAG y la ética profesional en la Contaduría Pública; Actualidad, priorizando publicaciones recientes (2022–2026); Rigor académico en las publicaciones recientes; Diversidad epistemológica, integrando enfoques éticos, sociotécnicos, normativos y educacionales. Lo anteriormente señalado, permitió un corpus equilibrado y representativo para el análisis cualitativo propuesto.

El análisis se desarrolló en tres momentos secuenciales: 1) Lectura exploratoria, orientada a identificar la pertinencia temática y el enfoque general de cada documento; 2) Lectura analítica, en la cual se registraron los aportes sustantivos, dilemas éticos y la crítica en la matriz de análisis; 3) Síntesis interpretativa y argumentativa, en donde los hallazgos obtenidos fueron contrastados con los supuestos teóricos, permitiendo la elaboración de los resultados y discusión crítica.

24 La presente reflexión se inspira en una comunicación personal con sistemas de inteligencia artificial generativa (Chat GPT, Open AI, del día 16 de enero del 2026), a partir del cual se elaboraron las ideas centrales que aquí se amplían y reinterpretan.

Los criterios de inclusión fueron: publicaciones recientes (2022 – 2026), relacionados con la IAG, la ética en la contaduría pública y el futuro del ejercicio de esta profesión, los documentos analizados cuentan con arbitraje académico o institucional.

Los criterios de exclusión fueron: publicaciones de carácter divulgativo sin sustento metodológico sólido, publicaciones duplicadas o sin referencia completa, estudios ajenos al ámbito de la IAG, a la contaduría pública y a la prospectiva del ejercicio de la profesión contable (economía digital – metaverso).

El diseño es no experimental y transversal y se fundamenta en el análisis y revisión documental cualitativo, con apoyo en un análisis de contenido temático, que permitió identificar categorías analíticas definidas previamente. El rigor del estudio se garantizó mediante: coherencia interna,[25] transparencia metodológica,[26] triangulación teórica,[27] Consistencia argumentativa.[28]

Los instrumentos de sistematización principales que se utilizaron fueron, una: Comparativa entre el Código de Ética Profesional del Contador Público (2024), la IFAC–IESBA (2024) y la UNESCO e IA (2022) (Tabla 1) y La matriz de análisis del corpus documental sobre ética, contaduría e IAG (Tabla 2)

Resultados y discusión

La incorporación de la IAG en la Contaduría Pública representa un avance importante en la enseñanza tradicional del ejercicio profesional. A diferencia de tecnologías previas orientadas a la automatización operativa, la IAG participa en procesos de análisis, interpretación normativa y generación de recomendaciones, lo que impacta en el juicio profesional del contador público. Este fenómeno obliga a replantear la relación entre técnica, ética y responsabilidad, pilares históricos de la profesión contable.

Desde la perspectiva de la ética profesional, la discusión no se centra en la legitimidad en el uso y la aplicación de la IAG, sino en las formas de utilizar y practicar la eticidad. Por lo que, el reto se encuentra inicialmente, en el cómo se toman decisiones profesionales contables mediadas por algoritmos: la integridad, objetividad, la responsabilidad social y

25 Asegurando la alineación entre objetivos, categorías analíticas y supuestos teóricos.

26 Al explicitar el proceso de selección y análisis del corpus documental.

27 Contrastando distintas perspectivas conceptuales y normativas.

28 Al vincular sistemáticamente los hallazgos documentales con la reflexión ética propuesta.

profesional ya no dependen del actuar humano directo, sino también de la capacidad del contador público para cuestionar los resultados generados con herramientas inteligentes.

Por lo que, delegar sin criticar funciones cognitivas a la IAG implica olvidarnos del juicio profesional, al hacer a un lado la deliberación ético – moral por la eficiencia algorítmica. Esto se debe a que los resultados que nos brinda la herramienta tecnológica están condicionados por datos de su entrenamiento, supuestos implícitos y decisiones de diseño que normalmente llegan a producir sesgos, errores o interpretaciones normativas incompletas.

Aun cuando la IAG apoye el análisis contable o fiscal, la decisión final y sus consecuencias sociales, económicas y legales continúan recayendo en el profesional. Por tanto, la discusión ética revela conflicto entre la eficiencia técnica, la responsabilidad moral, la rapidez, la reducción de costos y la optimización de procesos. Sin embargo, desde una perspectiva ética, la responsabilidad profesional no es transferible a la tecnología.

La ética profesional, por tanto, debe asumirse en la enseñanza universitaria como una competencia transversal, orientada no solo al cumplimiento normativo, sino a la toma de decisiones responsables en entornos digitales complejos.[29]

Por otra parte, en el análisis y discusión y de acuerdo con el (IMCP, 2024), entre sus lineamientos sustantivos se puede afirmar que: la responsabilidad del contador público, tanto legal como ética es indelegable;[30] El contador público cuando firma y responde: la IAG no asume responsabilidad.[31]

Es importante tener el debido cuidado y hacer verificación obligatoria, validando cálculos, citas normativas, supuestos, vigencia y coherencia contable-fiscal antes de usarlo. Y, no aceptar sin crítica resultados “persuasivos” de la IAG; aplicar contraste con fuentes primarias (ley, reglamento, NIF/NIIF, jurisprudencia aplicable, criterios, entre otros).

29 La literatura revisada en el corpus de este estudio sobre ética e inteligencia artificial —impulsada por organismos internacionales como la UNESCO— refuerza la idea de que el desarrollo tecnológico debe estar subordinado al interés público. En el caso de la Contaduría Pública, esto implica que la IAG debe concebirse como una herramienta al servicio del juicio ético profesional, y no como un sustituto de la deliberación ética que define la esencia de la profesión.

30 Toda salida de IAG usada en dictámenes, informes, papeles de trabajo o asesoría debe considerarse insumo, no decisión final.

31 Se debe de contar con transparencia profesional por uso de la IAG y declarar el uso de la misma cuando sea relevante para el usuario/cliente/autoridad (por ejemplo: elaboración de borradores, análisis preliminares, redacción de notas, clasificación). Además, evitar “presentar como certeza” lo que es probabilístico o no verificable (integridad).

Es necesario prohibir ingresar a la IAG información sensible identificable de clientes, RFC, estados financieros no públicos, estrategias fiscales internas, datos personales, sino existe control contractual/técnico (UNESCO – IESALC, 2023), Usar anonimización, minimización de datos y entornos autorizados (IMCP, 2024).

Procurar conservar evidencia mínima: objetivo del prompt, fecha, versión/herramienta, datos usados, salidas relevantes, validaciones realizadas y decisión final. Trazabilidad y evidencia -papeles de trabajo- (IFAC, 2023), además de contar con supervisión humana significativa, a través de un profesional competente que revise: supuestos, materialidad, riesgos,[32] cumplimiento normativo y comunicación al cliente. Además de evitar el uso de la IAG en decisiones de alto impacto.

Y, algo importante es el no usar a la IAG para fabricar evidencia, simular auditoría, "rellenar" papeles de trabajo o generar documentación que no corresponda a hechos verificables -riesgo directo de conducta indebida- (IMCP; 2024).

En relación con los supuestos teóricos que previamente fueron planteados en este estudio, éstos deben ser entendidos como proposiciones interpretativas que orientaron el análisis documental.

Supuesto teórico 1. Ética como eje articulador de la IAG en la educación contable

La incorporación de la IAG en la educación contable solo contribuye al fortalecimiento del aprendizaje profesional cuando se realiza con eticidad, integridad académica, responsabilidad social y juicio consciente y crítico. Se afirma también que la IAG no es neutral, por lo que, hay que prestar especial atención en la toma de decisiones éticas.

Supuesto teórico 2. Transformación pedagógica impulsada por la IAG

El uso de la IAG en la formación del contador público no produce efectos significativos si se reduce a la incorporación de herramientas tecnológicas, sino que exige un rediseño curricular y pedagógico orientado al desarrollo del pensamiento crítico y la evaluación auténtica de los aprendizajes. La efectividad formativa de la IAG depende, de su integración en modelos educativos centrados en competencias éticas y analíticas.

32 Contar con Comité u Oficial de Ética digital y un protocolo para incidentes (UNESCO – IESALC, 2023).

Supuesto teórico 3. Reconfiguración del perfil profesional del contador público

El uso creciente de la IAG redefine el perfil profesional del contador público, desplazando el énfasis de tares operativas hacia funciones de análisis, interpretación y toma de decisiones éticas. Por tal motivo, la formación contable debe de fortalecer el juicio profesional y la responsabilidad social como elementos fundamentales.

Supuesto teórico 4. Impacto social y confianza pública

La formación del contador público mediada por la IAG tiene implicaciones directas en la confianza social, la fe pública y la ciudadanía fiscal. Un uso éticamente responsable orientado por la IAG, contribuye al fortalecimiento de la legitimidad social de la profesión contable.

Los supuestos teóricos del estudio se presentan a nivel de resultados de la siguiente manera: 1) se asume que la ética profesional es un eje articulador para la incorporación de la IAG en la formación del contador público; 2) se afirma que la IAG solo genera valor educativo cuando se integra en modelos pedagógicos orientados al desarrollo del pensamiento crítico, la evaluación auténtica y la reflexión crítica; 3) la expansión de la IAG reconfigura el perfil profesional del contador público, fortaleciendo la necesidad de competencias analíticas, interpretativas y de juicio profesional, frente a la automatización.

Conclusiones

El análisis cualitativo y documental realizado permitió concluir que la incorporación de la IAG en la Contaduría Pública representa un cambio estructural en la formación universitaria y por ende en la práctica profesional, cuyos efectos trascienden el ámbito técnico para situarse en el núcleo ético del ejercicio profesional. La evidencia analizada confirma que la IAG no constituye una herramienta neutral, sino un sistema sociotécnico que incorpora decisiones humanas, sesgos algorítmicos y criterios implícitos que influyen de forma directa en los procesos contables, fiscales, financieros y de auditoría.

En este contexto, se reafirma que la ética profesional del contador público continúa siendo un eje insustituible, incluso en circunstancias de alta automatización. Ningún sistema de IAG puede sustituir el juicio moral, la deliberación ética ni la responsabilidad social que caracterizan al ejercicio profesional contable.

Asimismo, el estudio permitió identificar dilemas éticos emergentes asociados al uso de la IAG, entre los que destacan la atribución de responsabilidades por errores algorítmicos, la protección de datos sensibles, la opacidad de los sistemas inteligentes y los dilemas entre eficiencia técnica y responsabilidad ética. Estas situaciones evidencian que la automatización contable, lejos de soslayar la dimensión ética del ejercicio profesional, la reconfigura y hace que sea más compleja.

Otro hallazgo relevante es la constatación de que algunos de los códigos de ética profesional vigentes, si bien constituyen un referente normativo, resultan no suficientes para atender los desafíos éticos que plantea la IAG. Esta insuficiencia no implica la necesidad de sustituir dichos códigos, sino de actualizarlos y complementarlos con criterios éticos explícitos que consideren la mediación algorítmica en la toma de decisiones profesionales.

Los resultados también subrayan la urgencia de reconceptualizar la formación ética del estudiante de Contaduría Pública en contextos de IAG. La educación contable debe de manera sistemática incorporar el desarrollo del pensamiento crítico, la ética aplicada y comprender los alcances de las tecnologías inteligentes.

Por lo que, el verdadero valor del contador público en la era de la IAG no se encuentra en su capacidad para operar tecnologías avanzadas, sino en su competencia ética, crítica y reflexiva para integrarlas de manera socialmente responsable.

Finalmente, este trabajo concluye que la formación ética del contador público debe fortalecerse de manera transversal, integrando la reflexión sobre la IAG en los planes de estudio, la capacitación profesional continua y la regulación ética de la práctica contable. En este sentido, la IAG representa una oportunidad para revalorizar el papel ético del licenciado en Contaduría Pública. Al mismo tiempo, el estudio contribuye en la construcción de un marco teórico–conceptual que permite comprender la relación entre ética profesional, Contaduría Pública e IAG, abriendo líneas de investigación futuras en materia de regulación, formación profesional y evaluación ética del uso de herramientas tecnológicas inteligentes en la profesión contable.

Referencias

Ballantine, J., Boyce, Gordon, & Stoner, G. (2024). Una revisión crítica de la IA en la educación contable: amenazas y oportunidades. En *Perspectivas críticas sobre Contabilidad*. Elsevier, vol. 99(C). https://doi.org/10.1016/j.cpa.2023.102662

Barráez-Herrera, D. (2022). Metaversos en el Contexto de la Educación Virtual. *Revista Tecnológica-Educativa Docentes 2.0*, *13*(1), 11-19. https://doi.org/10.37843/rted.v13i1.300

Becerra Peña, D. L. (2023). Desafíos tributarios de la digitalización de la economía: una reflexión en torno al Proyecto BEPS. *Anáhuac Journal*, *23*(45), 55-74. https://revistas.anahuac.mx/index.php/the_anahuac_journal/article/view/2023

Bustamante Morán, N. Y., Andrade Santander, K. M., Velásquez Cambell, D. C., Edwin Javier Lema Cusquill, E. J., & León Vera, M. A. (2025). Percepción de los Docentes Sobre la Integración de la Inteligencia Artificial en los Procesos de Enseñanza y Aprendizaje. *Ciencia Latina Revista Científica Multidisciplinar*, *9*(2), 6482–6495. https://doi.org/10.37811/cl_rcm.v9i2.17382

Bolaño-García, M., & Duarte-Acosta, N. (2024). Una revisión sistemática del uso de la inteligencia artificial en la educación. *Revista Colombiana de Cirugía*, *39*(1), 51-63. https://doi. org/10.30944/20117582.2365

Camargo Arboleda, M. A., & Herrera Rocha, J. (2025). El impacto de la Inteligencia Artificial y la Gamificación en la Educación Contable. *ASCE Magazine*, *4*(1), 120–133. https://doi.org/10.70577/ASCE/120.133/2025

Chicaiza, V., Padilla R., Chicaiza, S., & Paredes, L. (2022). Tecnología de la realidad aumentada en el Interaprendizaje. *Revista científica mundo de la investigación y el conocimiento*. *6*(1), 145-155. https://recimundo.com/index.php/es/article/view/1514

IMCP (2024). *Código de Ética del Instituto Mexicano de Contadores Públicos*. IMCP.

Daza, Mercado, M. A. (2022). Ética y actitudes hacia el cumplimiento tributario en el Estado de Jalisco. *CIMEXUS Revista de Investigación Económica y Empresarial*, *17*(1), 89–108. https://doi.org/10.33110/cimexus170105

Daza Mercado, M. A., Aguilar Arámbula, F. J., Dávalos Abad., E. A., Sánchez Aldrete, J. L., Rodríguez Hinojosa, E., et al. (2006). *Ética profesional en la contaduría pública*. Universidad de Guadalajara. https://sitioanterior.cucea.udg.mx/sites/default/files/publicaciones/documentos/etica_contaduria01.pdf

Daza-Ramírez, M. T. (2025). Meta's privacy practices on Facebook: compliance, integrity, and a framework for excellence. *Discover Artificial Intelligence*. https://doi.org/10.1007/s44163-025-00388-5

Federación Internacional de Contadores (2023). *Handbook of the International Code of Ethics for Professional Accountants*. IFAC. https://ifacweb.blob.core.windows.net/publicfiles/2024-12/2023-Manual-2023-IESBA-Codigo-de-Etica-Spanish.pdf

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2021). AI People—An ethical framework for a good AI society. *Minds and Machines*, *31*(2), 1–29. https://doi.org/10.1007/s11023-021-09554-3

González, S. (2023). Ventajas y riesgos contables en el metaverso: lineamientos antifraude. *Revista de Auditoría y Control*, *12*(3), 201-220.

Organización para la Cooperación y el Desarrollo Económicos (2021). *Perspectivas de la Educación Digital de la OCDE 2021: Superando las Fronteras con Inteligencia Artificial, Blockchain y Robótica*. OCDE. https://intef.es/wp-content/uploads/2022/10/2022_01_Pushing-the-frontiers_INTEF.pdf

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura & Instituto Internacional para la Educación Superior en América Latina y el Caribe (2023). *ChatGPT e inteligencia artificial en la educación superior: Manual para las partes interesadas de la educación superior*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385146_spa

Palermo, C. (2024). *El impacto de la Inteligencia Artificial en la contabilidad empresarial*. Universidad de Palermo. https://www.palermo.edu/Archivos_content/contadorpublico/newsletter/2024/Enero/

Rodríguez García, B., Martínez Cortés, G. D., & Prieto Moreno, E. B. (2024). *Percepción de la inteligencia artificial (IA) en estudiantes de contaduría pública de la Corporación Universitaria Minuto de Dios* (UNIMINUTO), *Centro Universitario Villavicencio*. Repositorio Institucional UNIMINUTO. https://repository.uniminuto.edu/items/7dfbff9c-86c1-4047-abab-8f2092941711

Villalba Quinta, L. P. (2024). La inteligencia artificial en la contaduría pública: Transformación e innovación en la profesión contable. *Revista de Ciencias Empresariales, Tributarias, Comerciales y Administrativas*, *3*(2), 4–9.

Wanden -Berghe F. (2023). *Blockchain e inteligencia artificial en el sistema de información contable: la disrupción de la partida triple* [Tesis de grado, Universidad de Alicante] Repositorio Institucional. http://hdl.handle.net/10045/135180

06.

Integración de la inteligencia artificial en procesos de investigación en instituciones educativas en Zacatecas, México

Alejandro Guadalupe Rincón Castillo
Escuela Normal Rural Gral. Matías Ramos Santos, México
ORCID: 0000-0002-4905-9535

Cándida Marcela Rodríguez Chávez
Escuela Normal Rural Gral. Matías Ramos Santos, México
ORCID: 0000-0002-3213-1537

Luis Alonso Castañeda Negrete
Escuela Normal Rural Gral. Matías Ramos Santos, México
ORCID: 0000-0002-7417-0200

Daniel Alberto Mejía Herrera
Escuela Normal Rural Gral. Matías Ramos Santos, México
ORCID: 0009-0003-6731-4071

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.06

06.

Integración de la Inteligencia artificial en procesos de investigación en instituciones educativas de Zacatecas, México

Introducción

La integración de la Inteligencia Artificial (IA) en los procesos de investigación científica y educativa representa una de las transformaciones más relevantes del ámbito académico. En instituciones zacatecanas como la Escuela Normal Rural Gral. Matías Ramos Santos (ENRGMRS), México, y la Universidad Autónoma de Zacatecas (UAZ), México, surge la necesidad de examinar con detalle cómo se incorpora esta tecnología en las distintas etapas del trabajo de investigación científica, desde la revisión bibliográfica hasta la redacción y difusión de resultados. La primera institución se dedica principalmente a formar docentes en educación primaria, mientras que la segunda, se centra en la investigación multidisciplinaria. Esta diferencia facilita la identificación de similitudes y contrastes en el empleo de la IA, influenciados por el acceso a recursos tecnológicos, la preparación del personal docente y las directrices institucionales.

En este sentido, la literatura actual documenta de manera sistemática tanto las ventajas como los desafíos asociados al uso de la IA en la investigación. Ruiz (2024) explica que la automatización de tareas repetitivas libera tiempo para análisis más profundos, aunque reconoce retos éticos y prácticos importantes, como los sesgos algorítmicos que reproducen prejuicios de los datos de entrenamiento y comprometen la validez de los hallazgos. Asimismo, menciona preocupaciones por la privacidad de los participantes y el riesgo de exclusión de grupos subrepresentados, por lo que propone un enfoque integral que incluya marcos éticos sólidos, colaboración interdisciplinaria, formación especializada, transparencia algorítmica y cooperación internacional.

Por su parte, Díaz (2024) enfatiza que los investigadores deben dominar, no solo el manejo técnico de la IA, sino también su aplicación responsable y ética, considerando las repercusiones sociales de los resultados obtenidos. De manera complementaria, Acuña

(2025) constata que esta tecnología ya se emplea en etapas clave como la revisión de literatura, la planificación de proyectos y el procesamiento de grandes volúmenes de datos, pero advierte sobre riesgos éticos y sesgos que podrían distorsionar los resultados. En consecuencia, concluye que es indispensable capacitar a los investigadores educativos incorporando principios éticos y pensamiento crítico, de modo que la IA se perciba como apoyo y no como reemplazo de la capacidad humana.

A su vez, Suazo (2023) amplía el panorama al señalar que la IA genera riesgos éticos, legales y sociales que requieren regulación adecuada, entre ellos la vulneración de la confidencialidad de datos, la discriminación algorítmica, los impactos ambientales negativos y las consecuencias para el empleo, la educación y la autonomía humana. En línea con esta preocupación, Govanes y Lopezosa (2024) resaltan la importancia de una incorporación responsable mediante la colaboración entre personas y máquinas, apoyada en el pensamiento crítico como mecanismo de supervisión esencial.

De igual forma, Berrios y Márquez (2025) detectan un desconocimiento generalizado sobre las capacidades de los agentes inteligentes en contextos formativos para la investigación científica, lo que motiva el diseño y evaluación de estrategias pedagógicas que integren agentes generativos en la elaboración de proyectos. Su estudio confirma resultados positivos que evidencian mejoras en el aprendizaje investigativo. Molina (2024), por otro lado, analiza la evolución del concepto y su aporte a la educación y los procesos metodológicos, identificando desafíos agrupados en la relevancia para el cuarto objetivo de desarrollo sostenible, las limitaciones en la formación de docentes investigadores, las consideraciones éticas en la escritura científica y los problemas emergentes en el ámbito educativo.

En paralelo, Burgos et al. (2023) sostienen que todo el ciclo de investigación y redacción requiere supervisión constante de expertos humanos para garantizar exactitud, coherencia y credibilidad, por lo que recomiendan concebir la IA como colaborador ético y destacan la urgencia de aprender a manejarla. Juca (2023) aporta evidencia empírica al reportar que ningún docente identificó textos generados por IA, otorgándoles una Escala promedio de 8.88/10 y registrando solo 1% de similitud en herramientas antiplagio, lo que demuestra la dificultad actual para distinguirlos de producciones humanas.

Asimismo, Medina (2024) concluye que la IA tiene un potencial considerable para transformar la educación al fortalecer enseñanza, aprendizaje y gestión académica, aunque su éxito depende de marcos éticos institucionales, capacitación continua y una cultura de

adaptación al cambio. Vimos et al. (2024) encuentran, en un contexto concreto, que el 84% de los docentes participan en investigación científica y que el 72% considera que la la mejora la calidad de los textos académicos, facilitando procesos más eficientes.

Por su parte, Cedeño et al. (2023) revelan un consenso amplio sobre desafíos éticos en la investigación universitaria, destacando la opacidad algorítmica, los sesgos éticos y la necesidad de políticas institucionales claras. Finalmente, Román (2023) determina que herramientas como *ChatGPT* mejoran la calidad y eficiencia en la redacción de artículos científicos sin afectar la originalidad, por lo que los docentes recomiendan su uso en proyectos futuros.

Dos aspectos destacan al comparar las perspectivas de los autores revisados, en primer lugar, existe una coincidencia en reconocer los beneficios prácticos: automatización de tareas repetitivas, mejora en la eficiencia, apoyo en revisión bibliográfica, análisis de datos y redacción, como las preocupaciones éticas recurrentes: sesgos algorítmicos, falta de transparencia, privacidad de datos, riesgo de discriminación y necesidad de supervisión humana. Prácticamente todos los autores coinciden en que la IA debe entenderse como herramienta complementaria y no como sustituto, y en que su adopción responsable requiere formación, marcos éticos y regulación. En segundo lugar, se observa que, autores como Ruiz (2024), Suazo (2023) y Cedeño et al. (2023) profundizan en riesgos amplios (éticos, legales, sociales, ambientales y de discriminación), otros como Berrios y Márquez (2025), Acuña (2025) y Molina (2024) centran su atención en el ámbito educativo y pedagógico, destacando la formación docente, el aprendizaje investigativo y la alineación con objetivos de desarrollo sostenible. Asimismo, trabajos empíricos como los de Juca (2023), Vimos et al. (2024) y Román (2023) aportan datos cuantitativos sobre percepción, uso real y dificultad de detección, en contraste con las reflexiones más teóricas o propositivas de Díaz (2024), Govanes y Lopezosa (2024) y Burgos et al. (2023).

La literatura revisada presenta que la IA se aplica en etapas como la revisión de literatura, el diseño de proyectos, el análisis de datos a gran escala y la redacción de textos académicos. Ofrece mejoras en eficiencia y calidad, pero viene acompañada de preocupaciones persistentes sobre sesgos, ética, privacidad, transparencia y la necesidad de supervisión humana. Estas aportaciones conforman un marco teórico que reconoce tanto las posibilidades transformadoras como los riesgos que demandan regulación, formación y uso responsable.

En el contexto de la ENRGMRS y la UAZ se busca comprender con mayor precisión las particularidades de esta incorporación en entornos que pueden diferir en recursos, enfoques pedagógicos y acceso tecnológico.

El estudio se orienta por las siguientes preguntas de investigación: ¿Cómo se emplea la IA en las distintas etapas del proceso investigativo? ¿Se observan diferencias en su uso entre la Escuela Normal Rural Gral. Matías Ramos Santos y la Universidad Autónoma de Zacatecas?, estas preguntas guían el propósito principal del trabajo que es, identificar las formas en que se incorpora la IA en las diversas etapas de la investigación y determinar si existen diferencias relevantes entre las dos instituciones, contribuyendo así a una comprensión de su adopción.

Método de investigación

El presente estudio adoptó un enfoque comparativo para analizar las percepciones sobre la aceptabilidad del uso de la IA en las etapas del proceso investigativo en dos instituciones educativas zacatecanas: la ENRGMRS y la UAZ. Siguiendo el marco conceptual del método comparativo, definido como un procedimiento de comparación sistemática de casos con fines de generalización empírica y verificación de hipótesis (De León & de la Garza, 2013), se compararon las respuestas de participantes de ambas instituciones para identificar similitudes y diferencias en las actitudes hacia la IA. Este enfoque permitió contrastar contextos institucionales diferenciados, influenciados por factores como el acceso a recursos tecnológicos y la orientación pedagógica.

Diseño de la investigación

La investigación fue de tipo descriptiva-comparativa, con un enfoque cuantitativo y un diseño no experimental transversal. Se basó en la recolección de datos a través de un cuestionario estructurado para describir las percepciones de aceptabilidad de la IA en fases específicas del proceso investigativo (preliminar y planificación, ejecución y análisis, comunicación y difusión) y compararlas entre los dos grupos institucionales. El diseño comparativo se aplicó en un nivel descriptivo, centrado en la comparación de medias y distribuciones de respuestas, sin manipulación de variables independientes. Este enfoque fue adecuado para estudios educativos que buscan contrastar prácticas en entornos similares, pero con diferencias institucionales, permitiendo la identificación de patrones empíricos sin pretensiones causales (De León & de la Garza, 2013).

Participantes

La población objetivo comprendió investigadores y estudiantes en formación investigativa de la ENRGMRS y la UAZ, instituciones de educación superior con distinto enfoque: la primera orientada a la formación de docentes en educación primaria, y la segunda a la investigación multidisciplinaria. La muestra fue no probabilística por conveniencia, compuesta por 39 participantes (21 de la ENRGMR y 18 de la UAZ), reclutados a través de redes institucionales. Los participantes incluyeron roles como docentes que investigan (n=14), investigadores en formación (tesistas) (n=15), investigadores (n=7) y editores-revisores (n=4). Se garantizó el anonimato y el consentimiento informado, conforme a principios éticos de investigación educativa.

Instrumento

Se utilizó un cuestionario autoadministrado diseñado para este estudio, distribuido vía *Google Forms*. El instrumento constó de cuatro secciones: Sección A: Fase Preliminar y de Planificación, Sección B: Fase de Ejecución y Análisis, Sección C: Fase de Comunicación y Difusión. Cada ítem se evaluó en una escala Likert de 5 puntos (1 = Totalmente Inaceptable; 2 = Poco Aceptable; 3 = Neutral / Depende del Contexto; 4 = Aceptable; 5 = Totalmente Aceptable y Recomendable), con espacio opcional para comentarios breves. Sección D: Preguntas de Perfil y Perspectiva. El cuestionario fue validado por expertos en metodología educativa y probado en una prueba piloto con 5 participantes para asegurar claridad y fiabilidad (coeficiente alfa de Cronbach estimado en 0.85 para las secciones A-C).

Procedimiento

La recolección de datos se realizó entre enero y febrero de 2026, mediante el envío del enlace de Google Forms a redes académicas de la ENRGMR y la UAZ. Se incluyó una introducción explicando el propósito del estudio, el anonimato y el tiempo estimado (10-15 minutos). Las respuestas se exportaron automáticamente a una hoja de cálculo de Excel. El procedimiento siguió protocolos éticos, evitando cualquier coerción y asegurando la confidencialidad de los datos.

Análisis de datos

Los datos se analizaron utilizando software estadístico (*Excel* y *SPSS*). Se calcularon estadísticas descriptivas (medias, desviaciones estándar, frecuencias y porcentajes) para cada

ítem y sección, estratificados por institución. Para la comparación entre grupos (ENRGMRS y UAZ), se emplearon pruebas inferenciales no paramétricas (prueba U de Mann-Whitney) dada la escala ordinal y el tamaño muestral moderado, con un nivel de significancia de $p < 0.05$.

Resultados

Los resultados se presentaron en dos partes: las estadísticas descriptivas de las escalas de aceptabilidad del uso de IA generativa en las tres fases de la investigación (preliminar y planificación, ejecución y análisis, y comunicación y difusión), estratificadas por institución y las estadísticas inferenciales que compararon ambos grupos mediante la prueba U de Mann-Whitney. Las escalas Likert se realizaron en 5 puntos (1 = Totalmente Inaceptable, 2 = Poco Aceptable, 3 = Neutral / Depende del Contexto, 4 = Aceptable, 5 = Totalmente Aceptable y Recomendable). Las variables de perfil (Sección D) se resumieron con frecuencias y porcentajes. No se encontraron diferencias significativas entre instituciones (todos los $p > .05$), lo que indicó percepciones similares en general.

Estadísticas descriptivas

La Tabla 1 presenta las medias (M) y desviaciones estándar (DE) para cada ítem y sección, estratificadas por institución.

Tabla 1
Medias y desviaciones estándar de las escalas de aceptabilidad del uso de IA generativa por institución

Tarea	EN M(DE)	UAZ M(DE)
Sección A: Fase Preliminar y de Planificación		
A1: Generar ideas iniciales	3.95 (0.74)	3.67 (1.33)
A2: Realizar búsqueda bibliográfica preliminar	4.05 (0.92)	3.83 (1.04)
A3: Redactar borradores del marco teórico	3.00 (1.26)	2.94 (1.43)
A4: Formular preguntas de investigación	3.05 (1.20)	2.89 (1.41)
A5: Sugerir diseños metodológicos	3.29 (1.31)	3.72 (1.02)
A6: Redactar ítems iniciales de cuestionario	3.38 (1.12)	3.61 (0.98)
Media Sección A	3.45 (0.87)	3.44 (0.97)

Tabla 1

Medias y desviaciones estándar de las escalas de aceptabilidad del uso de IA generativa por institución

Tarea	EN M(DE)	UAZ M(DE)
Sección B: Fase de Ejecución y Análisis		
B1: Transcribir automáticamente entrevistas	3.81 (1.03)	3.89 (0.76)
B2: Análisis estadístico descriptivo básico	3.52 (0.81)	3.56 (0.70)
B3: Análisis estadístico inferencial complejo	3.14 (0.85)	3.22 (1.06)
B4: Codificar automáticamente datos cualitativos	3.62 (0.97)	3.22 (1.06)
B5: Interpretar resultados estadísticos	3.19 (0.93)	3.11 (1.23)
B6: Triangular hallazgos	3.52 (0.87)	3.44 (0.98)
Media Sección B	3.47 (0.64)	3.41 (0.65)
Sección C: Fase de Comunicación y Difusión		
C1: Corregir gramática, ortografía y estilo	3.81 (0.60)	3.78 (1.17)
C2: Redactar resumen ejecutivo	3.19 (1.21)	3.61 (1.24)
C3: Redactar secciones metodológicas	3.00 (1.18)	3.22 (1.06)
C4: Redactar sección de Discusión	2.71 (1.19)	2.89 (1.45)
C5: Redactar Conclusiones y Recomendaciones	2.67 (1.20)	2.78 (1.52)
C6: Formatear referencias bibliográficas	3.86 (0.91)	3.94 (1.06)
Media Sección C	3.21 (0.85)	3.37 (1.01)

Nota. EN = Escuela Normal; UAZ = Universidad Autónoma de Zacatecas. Las escalas se realizaron en una escala de 1 (Totalmente Inaceptable) a 5 (Totalmente Aceptable y Recomendable).

En términos generales, la aceptabilidad fue mayor en tareas auxiliares (p. ej., corrección gramatical, búsqueda bibliográfica preliminar) y menor en actividades centrales interpretativas (p. ej., redacción de discusiones o conclusiones). Los participantes de la ENRGMRS mostraron medias ligeramente superiores en la fase de ejecución (Sección B), mientras que los de UAZ presentaron mayor variabilidad en la fase de comunicación (Sección C).

Las Tablas 2, 3 y 4 muestra las frecuencias y porcentajes de las escalas por ítem, estratificadas por institución.

Tabla 2

Frecuencias y porcentajes de las escalas de aceptabilidad para tareas de la Sección A por institución

Tarea	Escala	EN f (%)	UAZ f (%)
A1	1	0 (0.0)	2 (11.1)
	2	0 (0.0)	1 (5.6)
	3	6 (28.6)	4 (22.2)
	4	10 (47.6)	5 (27.8)
	5	5 (23.8)	6 (33.3)
A2	1	0 (0.0)	1 (5.6)
	2	2 (9.5)	1 (5.6)
	3	2 (9.5)	2 (11.1)
	4	10 (47.6)	10 (55.6)
	5	7 (33.3)	4 (22.2)
A3	1	1 (4.8)	3 (16.7)
	2	9 (42.9)	5 (27.8)
	3	4 (19.0)	4 (22.2)
	4	3 (14.3)	2 (11.1)
	5	4 (19.0)	4 (22.2)
A4	1	2 (9.5)	3 (16.7)
	2	6 (28.6)	6 (33.3)
	3	4 (19.0)	2 (11.1)
	4	7 (33.3)	4 (22.2)
	5	2 (9.5)	3 (16.7)
A5	1	2 (9.5)	0 (0.0)
	2	5 (23.8)	1 (5.6)
	3	3 (14.3)	9 (50.0)
	4	7 (33.3)	2 (11.1)
	5	4 (19.0)	6 (33.3)

Tabla 2

Frecuencias y porcentajes de las escalas de aceptabilidad para tareas de la Sección A por institución

Tarea	Escala	EN f (%)	UAZ f (%)
A6	1	1 (4.8)	0 (0.0)
	2	4 (19.0)	2 (11.1)
	3	5 (23.8)	7 (38.9)
	4	8 (38.1)	5 (27.8)
	5	3 (14.3)	4 (22.2)

Nota. Escala: 1 = Totalmente Inaceptable; 2 = Poco Aceptable; 3 = Neutral/Depende del Contexto; 4 = Aceptable; 5 = Totalmente Aceptable y Recomendable. EN = Escuela Normal; UAZ = Universidad Autónoma de Zacatecas.

Tabla 3

Frecuencias y porcentajes de las escalas de aceptabilidad para tareas de la Sección B por institución

Tarea	Escala	EN f (%)	UAZ f (%)
B1	2	3 (14.3)	1 (5.6)
	3	4 (19.0)	3 (16.7)
	4	8 (38.1)	11 (61.1)
	5	6 (28.6)	3 (16.7)
B2	2	3 (14.3)	1 (5.6)
	3	5 (23.8)	7 (38.9)
	4	12 (57.1)	9 (50.0)
	5	1 (4.8)	1 (5.6)
B3	1	1 (4.8)	0 (0.0)
	2	3 (14.3)	6 (33.3)
	3	9 (42.9)	4 (22.2)
	4	8 (38.1)	6 (33.3)
	5	0 (0.0)	2 (11.1)
B4	1	0 (0.0)	1 (5.6)
	2	5 (23.8)	4 (22.2)
	3	0 (0.0)	4 (22.2)

Tabla 3

Frecuencias y porcentajes de las escalas de aceptabilidad para tareas de la Sección B por institución

Tarea	Escala	EN f (%)	UAZ f (%)
	4	14 (66.7)	8 (44.4)
	5	2 (9.5)	1 (5.6)
B5	1	0 (0.0)	2 (11.1)
	2	6 (28.6)	4 (22.2)
	3	6 (28.6)	4 (22.2)
	4	8 (38.1)	6 (33.3)
	5	1 (4.8)	2 (11.1)
B6	1	0 (0.0)	1 (5.6)
	2	3 (14.3)	1 (5.6)
	3	6 (28.6)	7 (38.9)
	4	10 (47.6)	7 (38.9)
	5	2 (9.5)	2 (11.1)

Nota. Escala: 1 = Totalmente Inaceptable; 2 = Poco Aceptable; 3 = Neutral/Depende del Contexto; 4 = Aceptable; 5 = Totalmente Aceptable y Recomendable. EN = Escuela Normal; UAZ = Universidad Autónoma de Zacatecas.

Tabla 4

Frecuencias y porcentajes de las escalas de aceptabilidad para tareas de la Sección C por institución

Tarea	Escala	EN f (%)	UAZ f (%)
C1	1	0 (0.0)	1 (5.6)
	2	1 (4.8)	2 (11.1)
	3	3 (14.3)	2 (11.1)
	4	16 (76.2)	8 (44.4)
	5	1 (4.8)	5 (27.8)
C2	1	2 (9.5)	1 (5.6)
	2	5 (23.8)	3 (16.7)
	3	3 (14.3)	3 (16.7)
	4	9 (42.9)	6 (33.3)

Tabla 4
Frecuencias y porcentajes de las escalas de aceptabilidad para tareas de la Sección C por institución

Tarea	Escala	EN f (%)	UAZ f (%)
	5	2 (9.5)	5 (27.8)
C3	1	2 (9.5)	0 (0.0)
	2	6 (28.6)	6 (33.3)
	3	5 (23.8)	4 (22.2)
	4	6 (28.6)	6 (33.3)
	5	2 (9.5)	2 (11.1)
C4	1	4 (19.0)	4 (22.2)
	2	5 (23.8)	4 (22.2)
	3	6 (28.6)	3 (16.7)
	4	5 (23.8)	4 (22.2)
	5	1 (4.8)	3 (16.7)
C5	1	4 (19.0)	5 (27.8)
	2	6 (28.6)	4 (22.2)
	3	5 (23.8)	2 (11.1)
	4	5 (23.8)	4 (22.2)
	5	1 (4.8)	3 (16.7)
C6	1	0 (0.0)	1 (5.6)
	2	2 (9.5)	1 (5.6)
	3	4 (19.0)	1 (5.6)
	4	10 (47.6)	10 (55.6)
	5	5 (23.8)	5 (27.8)

Nota. Escala: 1 = Totalmente Inaceptable; 2 = Poco Aceptable; 3 = Neutral/Depende del Contexto; 4 = Aceptable; 5 = Totalmente Aceptable y Recomendable. EN = Escuela Normal; UAZ = Universidad Autónoma de Zacatecas.

Estas distribuciones revelaron una tendencia general hacia escalas neutrales a aceptables (3–4), con menor frecuencia de escalas de aceptabilidad total (5) en las tareas centrales.

La Tabla 5 resume las características de perfil de los participantes (Sección D).

Tabla 5

Frecuencias y porcentajes de las características de perfil de los participantes por institución

Variable	Categoría	EN f (%)	UAZ f (%)
D1: Rol principal	Docente que investiga	9 (42.9)	5 (27.8)
	Investigador	7 (33.3)	3 (16.7)
	Investigador en formación (tesista)	4 (19.0)	9 (50.0)
	Editor-revisor	1 (4.8)	1 (5.6)
D2: Paradigma de investigación	Cualitativo	15 (71.4)	6 (33.3)
	Mixto	6 (28.6)	8 (44.4)
	Cuantitativo	0 (0.0)	4 (22.2)
D3: Postura hacia la IA generativa	Pragmática (útil con reglas claras)	15 (71.4)	7 (38.9)
	Precavida (riesgos superan beneficios)	2 (9.5)	6 (33.3)
	Entusiasta (herramienta revolucionaria)	1 (4.8)	3 (16.7)
	Crítica (compromete esencia de la investigación)	3 (14.3)	2 (11.1)
D4: Principal riesgo	Pérdida de pensamiento crítico y autonomía	14 (66.7)	13 (72.2)
	Producción de contenido plagiado/ no original	3 (14.3)	1 (5.6)
	Amplificación de sesgos en datos de entrenamiento	2 (9.5)	2 (11.1)
	Falta de transparencia y dificultad para auditar	2 (9.5)	1 (5.6)
	Otro	0 (0.0)	1 (5.6)

Nota. EN = Escuela Normal; UAZ = Universidad Autónoma de Zacatecas.

La mayoría se identificó como investigadores cualitativos o mixtos, con una postura pragmática hacia la IA (herramienta útil con reglas claras). El riesgo principal percibido fue la pérdida de pensamiento crítico y autonomía intelectual (≈70 % en ambos grupos).

Estadísticas inferenciales

La Tabla 6 muestra los resultados de las pruebas U de Mann-Whitney para comparar las escalas de aceptabilidad entre EN y UAZ.

Tabla 6

Resultados de la prueba U de Mann-Whitney para comparar las escalas de aceptabilidad entre instituciones

Tarea/Sección	U	P
Sección A		
A1	200.0	.756
A2	211.0	.509
A3	194.0	.896
A4	203.0	.695
A5	156.0	.345
A6	171.5	.619
Media Sección A	199.0	.788
Sección B		
B1	186.5	.952
B2	191.5	.950
B3	185.0	.917
B4	232.0	.183
B5	193.0	.918
B6	196.5	.833
Media Sección B	200.5	.755
Sección C		
C1	173.5	.628
C2	151.0	.272
C3	170.0	.589

Tabla 6

Resultados de la prueba U de Mann-Whitney para comparar las escalas de aceptabilidad entre instituciones

Tarea/Sección	U	P
C4	176.5	.729
C5	183.5	.885
C6	170.5	.582
Media Sección C	177.0	.745

Nota. Pruebas bilaterales. No se encontraron diferencias significativas ($p < .05$).

No se observaron diferencias significativas en ningún ítem ni en las medias de sección ($p > .05$), lo que sugirió que la afiliación institucional no influyó en las percepciones sobre la aceptabilidad de la IA generativa en tareas de investigación educativa.

Como se observa los participantes de ambas instituciones mostraron una aceptabilidad moderada hacia el uso de IA generativa en tareas de apoyo, como la generación de ideas preliminares (M global = 3.82) y la transcripción automática (M global = 3.85), pero expresaron reservas importantes en actividades que requieren alto nivel intelectual, como la interpretación de resultados (M global = 3.15) o la redacción de discusiones (M global = 2.79). Este patrón fue coherente con la preocupación predominante por la pérdida de autonomía intelectual, el riesgo más frecuentemente señalado. La ausencia de diferencias entre grupos sugirió que estas actitudes fueron ampliamente compartidas entre investigadores educativos en México, independientemente del tipo de institución, posiblemente reflejando valores éticos y pedagógicos comunes en el campo. Se recomendó realizar estudios longitudinales para monitorear posibles cambios en estas percepciones con el avance y mayor integración de las herramientas de IA generativa.

Discusión y conclusiones

Los resultados revelan una aceptación moderada de la IA generativa en las fases de la investigación educativa, con medias generales que oscilan entre neutrales y aceptables (Sección A: M = 3.45; Sección B: M = 3.44; Sección C: M = 3.29), y una mayor reticencia en tareas interpretativas centrales. Estas percepciones se alinean con la literatura revisada, que enfatiza el potencial de la IA para optimizar procesos auxiliares, pero advierte sobre riesgos éticos y cognitivos en etapas críticas. No obstante, la ausencia de diferencias significativas entre instituciones sugiere que estas actitudes podrían ser generalizables en otros contextos educativos.

En la fase preliminar y de planificación (Sección A), la alta aceptabilidad para tareas como la generación de ideas iniciales (M = 3.82 global) y búsquedas bibliográficas (M = 3.95 global) coincide con hallazgos de Ruiz Muñoz (2024), quien describe cómo la IA transforma la metodología al identificar patrones complejos y automatizar revisiones preliminares, liberando tiempo para análisis profundos. De manera similar, Díaz Subieta (2024) resalta el uso de la IA en revisiones sistemáticas para procesar grandes volúmenes de datos, lo que explica la percepción positiva en ítems como A2 (M = 3.95 en EN). Sin embargo, la menor aceptación en la redacción de borradores teóricos (M = 2.97 global) y formulación de preguntas (M = 2.97 global) refleja preocupaciones éticas destacadas por Suazo Galdames (2023), como sesgos algorítmicos que podrían perpetuar prejuicios en datos de entrenamiento, afectando la validez en investigaciones educativas. Esta reticencia también se observa en Acuña-Gamboa (2025), quien advierte sobre riesgos en la revisión de literatura, donde la IA podría generar sesgos si no se supervisa críticamente.

En la fase de ejecución y análisis (Sección B), la aceptación moderada para transcripciones automáticas (M = 3.85 global) y análisis descriptivos (M = 3.54 global) se alinea con la visión de Medina Romero (2024), quien propone aplicaciones de IA para analizar datos a gran escala en educación superior, mejorando la eficiencia en tareas repetitivas. Román Acosta (2023) corrobora esto al demostrar que herramientas como ChatGPT mejoran la calidad y eficiencia en redacción académica sin comprometer la originalidad, lo que explica la neutralidad en codificación cualitativa (M = 3.44 global). No obstante, la baja aceptabilidad en interpretación de resultados (M = 3.15 global) y triangulación (M = 3.49 global) resuena con las advertencias de Molina-Isaza (2024) sobre la pérdida de autonomía intelectual en investigación educativa, y con Burgos et al. (2023), quienes enfatizan que la IA, aunque útil para brainstorming, no sustituye el juicio humano en análisis inferenciales complejos. Este patrón subraya el riesgo principal identificado en D4 (pérdida de pensamiento crítico, 69% global), congruente con Goyanes y Lopezosa (2024), que en revisiones sobre ChatGPT en ciencias sociales destacan la necesidad de supervisión humana para evitar sesgos en métodos cualitativos y cuantitativos.

Para la fase de comunicación y difusión (Sección C), la alta aceptación en corrección gramatical (M = 3.80 global) y formateo de referencias (M = 3.90 global) es consistente con Vimos-Buenaño et al. (2024), quienes reportan que el 72% de docentes universitarios creen que la IA mejora la calidad de textos académicos. Berrios Zepeda y Márquez Mora (2024) apoyan esto al mostrar mejoras en redacción asistida por IA generativa en aulas, similar a la

neutralidad en resúmenes (M = 3.38 global). Sin embargo, la menor aceptación en discusiones (M = 2.79 global) y conclusiones (M = 2.72 global) refleja las preocupaciones éticas de Ruiz Muñoz (2024) sobre privacidad y exclusión de grupos subrepresentados, y de Suazo Galdames (2023) sobre impactos en la autonomía humana. Esto se vincula con la postura pragmática predominante en D3 (56% global), que aboga por reglas claras, alineada con recomendaciones de Díaz Subieta (2024) para marcos éticos en usos emergentes de IA.

En cuanto al perfil, la identificación mayoritaria con paradigmas cualitativos (54% global) y roles de docente-investigador (36% global) explica la precaución hacia la IA en fases interpretativas, como nota Acuña-Gamboa (2025) en contextos mexicanos. La falta de diferencias institucionales (todos $p > .05$) sugiere homogeneidad en percepciones, posiblemente por exposiciones similares a herramientas como ChatGPT, como en Burgos et al. (2023). No obstante, discrepancias con literatura más optimista (Medina Romero, 2024) podrían deberse a muestras limitadas a docentes educativos, donde preocupaciones pedagógicas por el pensamiento crítico son más pronunciadas.

Estos hallazgos implican la necesidad de marcos éticos holísticos, como propone Ruiz Muñoz (2024), fomentando colaboraciones interdisciplinarias para integrar IA responsablemente. Futuras investigaciones podrían explorar intervenciones longitudinales para mitigar riesgos, asegurando que la IA complemente, no reemplace, la rigurosidad humana en investigación educativa.

Respecto a las diferencias institucionales, no se detectaron variaciones significativas en las escalas de aceptabilidad (todos $p > .05$ en las pruebas U de Mann-Whitney), lo que sugiere que las actitudes hacia la IA son homogéneas entre ambas instituciones, posiblemente influenciadas por factores comunes como el acceso limitado a recursos tecnológicos en entornos rurales y universitarios zacatecanos, así como por valores éticos compartidos en la formación docente e investigativa. Esta similitud subraya que, independientemente del enfoque pedagógico, los investigadores educativos en Zacatecas perciben la IA como una herramienta pragmática (postura dominante en el 56% global), pero con necesidad de reglas claras para mitigar riesgos como sesgos algorítmicos, falta de transparencia y amplificación de prejuicios.

Las implicaciones de estos resultados son tanto teóricas como prácticas. Teóricamente, contribuyen al marco conceptual sobre la integración ética de la IA en la investigación educativa, reforzando argumentos de autores como Ruiz (2024) y Suazo (2023) sobre la

necesidad de marcos éticos sólidos y colaboración interdisciplinaria. Prácticamente, sugieren la urgencia de capacitar a docentes e investigadores en principios éticos y pensamiento crítico, alineados con objetivos de desarrollo sostenible como el ODS 4 (educación de calidad), para fomentar una adopción responsable que evite la exclusión de grupos subrepresentados y preserve la integridad académica. En el contexto zacatecano, esto podría traducirse en políticas institucionales que promuevan talleres de formación y guías de uso ético, adaptadas a las limitaciones locales en acceso tecnológico.

No obstante, el estudio presenta limitaciones inherentes a su diseño: la muestra no probabilística por conveniencia (n = 39) limita la generalización a otras instituciones, y el enfoque cuantitativo transversal no captura cambios longitudinales en percepciones ante el rápido avance de la IA. Además, la ausencia de análisis cualitativo profundo en los comentarios opcionales podría haber enriquecido la interpretación de contextos dependientes.

Para investigaciones futuras, se recomienda explorar diseños mixtos que incorporen datos cualitativos detallados, estudios longitudinales para monitorear evoluciones en la adopción de IA (por ejemplo, post-implementación de herramientas como *ChatGPT*), y comparaciones con instituciones de otros estados mexicanos para identificar influencias regionales. Asimismo, evaluar intervenciones pedagógicas específicas, como las propuestas por Berrios y Márquez (2025), podría proporcionar evidencia sobre estrategias para maximizar beneficios mientras se mitigan riesgos.

Referencias

Acuña-Gamboa, L. A. (2025). Inteligencia Artificial e investigación educativa en los contextos internacional y mexicano: Revisión de la literatura. *Revista Paraguaya de Educación a Distancia*. https://doi.org/10.56152/reped2025-vol6num2-art9

Berrios Zepeda, R., & Márquez Mora, L. (2025). Agente de Inteligencia Artificial Generativa en investigación científica. Un análisis explicativo del aprendizaje en el aula. *RIED-Revista Iberoamericana de Educación a Distancia*. https://doi.org/10.5944/ried.28.2.43545

Burgos, L. M., Suárez, L. L., & Benzadón, M. (2023). *Inteligencia Artificial ChatGPT y su utilidad en la investigación: El futuro ya está aquí*. Medicina (Buenos Aires). https://www.medicinabuenosaires.com/revistas/vol83-23/n3/499.pdf

Cedeño Meza, J. G., Maitta Rosado, I. S., Vélez Zambrano, M. L., & Palomeque Zambrano, J. Y. (2023). Investigación universitaria con Inteligencia Artificial. *Revista Venezolana de Gerencia*. https://doi.org/10.52080/rvgv26n94.14

De León, C., & de la Garza, E. (2013). *Método comparativo*. Universidad Autónoma de Nuevo León. http://eprints.uanl.mx/9802/1/Estudio%20Comparado.pdf

Díaz Subieta, L. B. (2024). El uso de la Inteligencia Artificial en la investigación científica. *Revista Historia de la Educación Latinoamericana*. https://doi.org/10.19053/uptc.01227238.18014

Goyanes, M., & Lopezosa, C. (2024). *ChatGPT en Ciencias Sociales: revisión de la literatura sobre el uso de Inteligencia Artificial (IA) de OpenAI en investigación cualitativa y cuantitativa*. Anuario ThinkEPI. https://thinkepi.scimagoepi.com/index.php/ThinkEPI/article/download/91634/66391/314943

Juca Maldonado, F. (2023). El impacto de la Inteligencia Artificial en los trabajos académicos y de investigación. *Revista Metropolitana de Ciencias Aplicadas*. https://www.redalyc.org/pdf/7217/721778121031.pdf

Medina Romero, M. Á. (2024). Aplicaciones de la IA para la investigación y la innovación en la educación superior. *Revista Social Fronteriza*. https://doi.org/10.59814/resofro.2024.4(4)336

Molina-Isaza, L. E. (2024). Aspectos determinantes de la IA en la investigación educativa. *Praxis*. https://doi.org/10.21676/23897856.6005

Román Acosta, D. (2023). Aplicación de la IA en la investigación académica: caso ChatGPT. *Revista Finanzas y Negocios*. https://revistas.ulatina.edu.pa/index.php/Finanzasynegocios/article/download/323/370/1333

Ruiz Muñoz, G. F. (2024). Implicaciones de la IA en la metodología de investigación. *RITI Journal*. https://doi.org/10.36825/RITI.12.26.003

Suazo Galdames, I. (2023). *IA en investigación científica*. SciComm Report. https://doi.org/10.32457/scr.v3i1.2149

Vimos-Buenaño, K. E., Viteri-Ojeda, J. C., Naranjo-Sánchez, M. J., & Novillo-Heredia, K. H. (2024). Uso de la IA en los procesos de investigación científica, por parte de los docentes universitarios. *Journal of Economic and Social Science Research*. https://doi.org/10.55813/gaea/jessr/v4/n4/143

07.

¿Pueden las máquinas salvar al maíz? Aprendizaje profundo para la detección de plagas

Antonio Juan Capistran-Abundez
Centro Nacional de Investigación y Desarrollo Tecnológico, México
ORCID: 0000-0001-8110-989X

Vitervo López-Caballero
Centro Nacional de Investigación y Desarrollo Tecnológico, México
ORCID: 0000-0002-1942-9558

Lucia Morales-Morales
Centro Nacional de Investigación y Desarrollo Tecnológico, México
ORCID: 0009-0006-6593-4762

Andrea Sánchez-Ruiz
Centro Nacional de Investigación y Desarrollo Tecnológico, México
ORCID: 0009-0005-7773-0575

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.07

07.

¿Pueden las máquinas salvar al maíz? Aprendizaje profundo en la detección de plagas

Introducción

Según la Organización de las Naciones Unidas (ONU), la Agenda 2030 para el Desarrollo Sostenible, es un plan de colaboración entre países interesados por garantizar la paz universal, la prosperidad, la libertad y los derechos humanos para guiar al planeta Tierra hacia la sostenibilidad. Con una visión integral del futuro de la humanidad, la propuesta tiene 17 Objetivos para el Desarrollo Sostenible (ODS), los cuales proporcionan una hoja de ruta para impulsar el crecimiento económico, social, cultural y la protección del medio ambiente.

La presente investigación se alinea con el ODS 2 que, de acuerdo con Huck (2023), promueve el incremento de la productividad agrícola y la sostenibilidad de los sistemas alimentarios. También se adscribe con el ODS 12, el cual fomenta el uso eficiente de los recursos naturales y la reducción del impacto ambiental. Además, Mejía (2003) menciona que la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), considera al maíz como una fuente de vida para quienes lo cultivan y por su adaptabilidad a diferentes climas, es uno de los granos alimenticios de mayor producción y consumo en el mundo.

De igual forma, la Secretaría de Agricultura y Desarrollo Rural de México (SADER), informa que el maíz es un alimento básico cultivable en los 32 estados del país, siendo así un componente fundamental del patrimonio gastronómico y cultural de la sociedad mexicana. Sin embargo, Santhi et al. (2024) reportan que el daño de plagas en diversas secciones del maíz, disminuye la producción. Por lo cual, Liakos et al. (2018) consideran adecuado vigilar la salud de los plantíos durante el crecimiento para que el rendimiento agrícola aumente. Mientras que, Barbedo (2018) menciona que la inspección visual de plagas en los cultivos puede originar sesgos. Por lo tanto, se considera a la agricultura inteligente como un recurso

recomendable para abordar este desafío. Es decir, según Kamilaris et al. (2017) la agricultura asociada con la tecnología, mejora los recursos obtenidos de la actividad agraria.

Finalmente, Yu et al. (2021) mencionan que a partir de la integración de la inteligencia artificial en actividades agrícolas, diversos estudios abordan la clasificación de plagas en plantas de arroz, mango, manzana, pepino, té y otros cultivos. También, Xu et al. (2023) considera que, la clasificación de imágenes tiene diversas aplicaciones en la agricultura. Por lo tanto, el objetivo de este trabajo es reportar un conjunto de artículos en los cuales se investigan algoritmos de automatización para detectar plagas en imágenes de hojas de maíz capturadas con dron y celular.

Descripción del problema de investigación

Según, Brewbaker & Hawaii (2003) el maíz es fundamental para las finanzas del mundo debido a que la solicitud del cultivo aumenta conforme crece la población mundial. Sin embargo, Liakos et al. (2018) consideran que las plagas son un desafío constante para los agricultores debido a las complicaciones para controlarlas tanto en condiciones de cultivo al aire libre como en invernadero y según, L. Paliwal (2001), destacan por ser las más visibles, las enfermedades foliares tales como: los tizones, las manchas y las royas, ya que a primera vista son alarmantes.

Con base en lo anterior, Qian et al. (2022) recomiendan identificar las plagas en etapas iniciales para evitar la propagación al resto del cultivo, reducir el uso de pesticidas y mejorar la cantidad de producción. Sin embargo, Wahyuningrum et al. (2023) mencionan que la detección manual es difícil y propensa a errores en entornos de campo. Por tal razón, Yu et al (2021) lo consideran un desafío constante para los agricultores detectar plagas durante la etapa de crecimiento del maíz.

Por su parte, Rouhiainen (2020) menciona que la integración de la inteligencia artificial a la agricultura, permite a los productores tomar mejores decisiones para incrementar la producción. Por lo tanto, el objetivo general de la presente investigación es desarrollar e implementar un sistema basado en aprendizaje profundo mediante el análisis de imágenes capturadas con dron y celular para la detección de plagas en cultivos de maíz. Para ello, se proponen los siguientes objetivos específicos:

1. Diseñar un conjunto de imágenes capturadas con dron y celular.

2. Entrenar un modelo de aprendizaje profundo con técnicas de procesamiento de imágenes.

3. Validar el modelo de aprendizaje profundo realizando pruebas con el conjunto de imágenes.

4. Desarrollar una plataforma web para identificar plagas en imágenes de cultivos de maíz.

Justificación de la investigación

Según la FAO, se reconoce al maíz a nivel mundial por su adaptabilidad a diferentes condiciones climáticas y en México es pilar de la agricultura debido a que se cultiva en los 32 estados del país siendo así un grano estratégico por su valor económico y pertinente para la seguridad alimentaria. Sin embargo, según Wahyuningrum et al. (2023), la presencia de enfermedades en las hojas de maíz son un riesgo constante para la producción, lo cual disminuye la disponibilidad del cereal.

Por su parte, Xu et al. (2023) mencionan que tradicionalmente, la detección de plagas es una actividad realizada mediante inspección visual, lo cual consume tiempo, es propensa a errores y difícil de realizar en campos agrícolas grandes, lo cual puede llevar a daños irreparables en los cultivos, menor producción y afectación negativa económica de los agricultores.

Con base en lo anterior, es recomendable implementar sistemas automáticos que ayuden a los productores de maíz a detectar las plagas en etapas iniciales de infestación para que les sea posible tomar medidas preventivas. En este sentido el uso de redes neuronales convolucionales se plantea como una solución innovadora ya que estas herramientas de aprendizaje profundo tienen el potencial de analizar grandes conjuntos de imágenes y detectar patrones de interés. Además, estas técnicas permiten automatizar la detección de plagas, lo cual puede reducir tiempos, costos operativos y minimizar el uso de insumos agrícolas como pesticidas, contribuyendo así a la sostenibilidad ambiental. En este sentido, para abordar uno de los desafíos permanentes de la agricultura, la presente investigación tiene el propósito de detectar plagas en cultivos de maíz utilizando técnicas de aprendizaje profundo con redes neuronales convolucionales, las cuales tienen buenos resultados para el reconocimiento de imágenes.

Marco conceptual

Inteligencia artificial

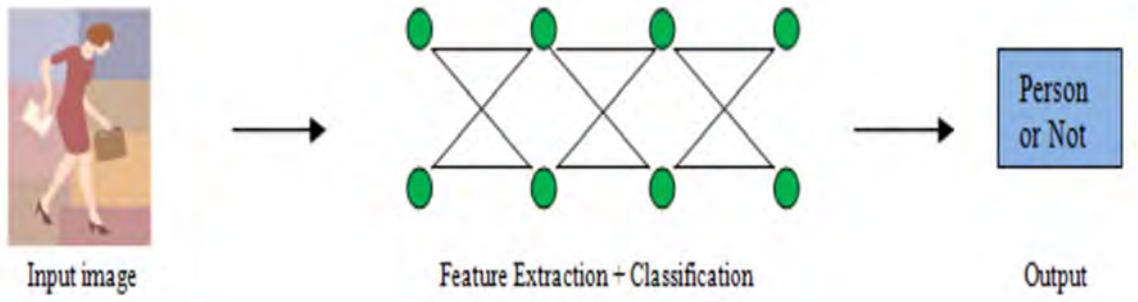
Según Russell y Norvig (2021), es el estudio de agentes que reciben percepciones del entorno y realizan acciones. Esta perspectiva se centra en el concepto de agente inteligente, que asigna secuencias de percepción a acciones específicas. Esta definición se alinea con el enfoque de "sistemas que actúan racionalmente", enfatizando la creación de sistemas que aprenden qué maximizar y funcionan bajo la incertidumbre de los objetivos humanos.

Aprendizaje profundo

Según Hendrycks (2024), son un conjunto de técnicas que utiliza redes neuronales con múltiples capas para crear modelos que pueden aprender de grandes conjuntos de datos. Las redes neuronales utilizan capas de nodos interconectados para transformar las entradas en salidas. Por ejemplo, una red neuronal profunda entrenada para el reconocimiento de objetos en imágenes puede aprender a identificar patrones en los datos de píxeles sin procesar, tal como se observa, de forma general el funcionamiento del aprendizaje profundo en la Figura 1.

Figura 1

Funcionamiento general del aprendizaje profundo



Nota. Tomada de Krishna y Kalluri (2019).

Red neuronal convolucional

Según Courville (2016), es una red que emplea una operación matemática llamada convolución, la cual es un tipo especializado de operación lineal. Se especializa en procesar

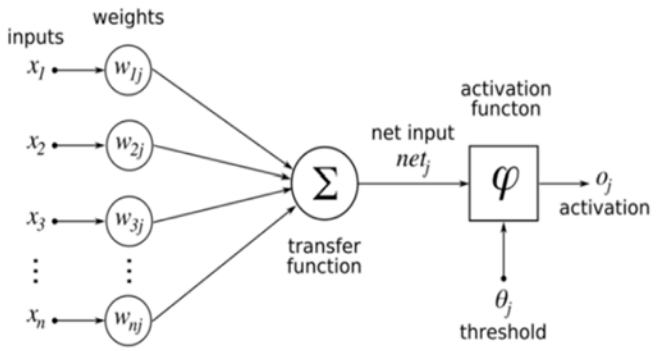
datos que tienen una topología conocida similar a una cuadrícula. Es decir, mediante la aplicación de múltiples capas de convolución, seguidas de capas de activación, agrupamiento y normalización, pueden modelar relaciones complejas entre los elementos de una imagen, permitiendo aplicaciones avanzadas como la detección de objetos y la segmentación semántica.

Funciones de las redes neuronales convolucionales

Se estructuran en capas sucesivas que permiten la extracción automática de características de las imágenes. El proceso inicia con las capas convolucionales, aplicando filtros para detectar patrones como bordes, texturas y formas. Estas capas permiten a la red identificar características locales en las imágenes, que luego se combinan a medida que avanzan a través de las capas. Después, las capas de agrupamiento reducen la resolución de las imágenes para mantener las características más importantes y disminuir la complejidad computacional. Finalmente, las capas totalmente conectadas integran la información extraída por las capas anteriores y realizan la clasificación. En la Figura 2, se observa de forma general el funcionamiento de una red neuronal.

Figura 2

Representación esquemática de una neurona artificial



Nota. Tomada de UNAL (2020).

Redes neuronales residuales (ResNet)

Según, He et al. (2016) son arquitecturas de aprendizaje profundo que utilizan conexiones de atajo basadas en la identidad dentro de los bloques de la red. A diferencia de las redes de propagación hacia adelante simples, las ResNet incorporan estas conexiones para facilitar el entrenamiento de modelos muy profundos, mejorar el flujo de información y

proporcionar un mejor espacio de funciones. Es decir, la entrada de la capa anterior se suma a la salida de la capa actual por lo cual, esta conexión residual facilita el aprendizaje de la red y mejora su rendimiento en actividades de clasificación de imágenes, detección de objetos, segmentación semántica, extracción de características y transferencia de aprendizaje.

Como ejemplo, la ResNet-50 destaca por ser una arquitectura de gran profundidad cuya principal innovación es el uso de bloques residuales tipo “cuello de botella”. Su estructura interna se organiza en una secuencia de 16 bloques residuales distribuidos en grupos de 3, 4, 6 y 3, donde cada bloque integra tres capas convolucionales con dimensiones de 1x1, 3x3 y 1x1 pixeles.

Maíz

Según, L. Paliwal (2001) es de los granos alimenticios más antiguos con origen en México de la familia Poáceas o Gramíneas, tribu *Maydeas*, *Zea mays*. El maíz no crece en forma silvestre y no puede sobrevivir en la naturaleza, por lo cual es una planta completamente domesticada. Además, es uno de los cereales de mayor consumo y una de las especies vegetales más productiva.

Plagas del maíz

Según, Savary et al. (2019), son organismos bióticos principalmente insectos, hongos, bacterias y virus que interactúan negativamente con el cultivo, provocando daños fisiológicos, estructurales o funcionales causantes de pérdidas de rendimiento, calidad del grano y rentabilidad económica, siendo ejemplo, las plagas del cogollo que dañan los extremos de los brotes de tallos y raíces afectando el crecimiento vertical de la planta.

Trabajos relacionados

Con base en la descripción del problema, la búsqueda de investigaciones afines se realiza en idioma inglés durante el periodo de años 2018-2025 en los buscadores académicos, *Google Scholar*, *IEEE Xplore*, *Multidisciplinary Digital Publishing Institute* (MDPI) y *ScienceDirect* a partir de cadenas de búsqueda formadas por las siguientes palabas: “deep learning algorithms”, “pest detection in corn crops” y “drone and cell phone images”. Se observa que, en los artículos revisados, se abordan tres enfoques de algoritmos de inteligencia artificial: aprendizaje automático, aprendizaje profundo y aprendizaje hibrido, los cuales se describen a continuación.

Enfoque con Aprendizaje Automático

El propósito de este enfoque es abordar investigaciones que han implementado algoritmos de aprendizaje automático, tales como: *Support Vector Machine* (SVM), *Multinomial Naïve Bayes* (MNB), *K-Nearest Neighbor* (KNN) y *Decision Trees* (DT), orientados a la identificación de plagas y enfermedades foliares en imágenes de hojas de maíz. La eficacia de estos modelos se evalúa mediante métricas de desempeño como la precisión, la matriz de confusión y la sensibilidad.

En este sentido, Barbedo (2018) implementó un modelo de aprendizaje por transferencia utilizando la red neuronal GoogLeNet con la Neural Network Toolbox de MATLAB 2017b. Se entrenó y validó con un conjunto de 50,000 imágenes, orientadas a la detección de las principales enfermedades del maíz. Por otra parte, Aravind et al. (2018) proponen un enfoque basado en el procesamiento de imágenes, empleando un modelo de máquina de vectores de soporte multiclase para analizar 2,000 imágenes de PlantVillage, las cuales fueron clasificadas en mancha de cercospora, óxido, pudrición de hojas y hojas sanas alcanzando un 83.7% de precisión.

De forma parecida, en el estudio de Resti et al. (2022) se analizaron 761 imágenes proporcionadas por la Universidad de Sriwijaya para identificar las enfermedades óxido, mildiu y mancha foliar así como las plagas *spodoptera frugiperda* y *heliothis armigera* en plantas de maíz. En este caso, se aplicaron los algoritmos *Multinomial Naïve Bayes* y *K-Nearest Neighbor*, con resultados de precisión del 92.72% y 99.54%, respectivamente.

Finalmente, en la investigación de Idress et al. (2024) se analizaron 600 imágenes de *PlantVillage*, previamente segmentadas mediante el método *K-means clustering* y caracterizadas a través de medidas estadísticas como la media y la desviación estándar. Posteriormente, se aplicaron técnicas de clasificación como árboles de decisión, *K-Nearest Neighbor*, *Support Vector Machine* (SVM) y Redes Neuronales Artificiales (ANN), logrando un 92.7% de precisión con ANN y un 91.3% con SVM en la clasificación de mancha foliar, roya común y hojas sanas.

Enfoque con aprendizaje profundo

Debido a los buenos resultados de las técnicas de aprendizaje profundo, diversas investigaciones recientes coinciden en la aplicación de redes neuronales convolucionales para identificar plagas en imágenes de hojas de maíz. De forma general, se han utilizado

conjuntos de imágenes públicas como *PlantVillage*, repositorios académicos de libre acceso, colecciones locales obtenidas en campo e incluso imágenes sintéticas generadas mediante técnicas de aumento de datos, las cuales se validan con métricas ampliamente aceptadas en la literatura tales como precisión, sensibilidad, exactitud y puntaje F1, lo que ha permitido observar la eficacia de estos modelos.

En una primera línea de investigación, varios estudios prefieren las redes clásicas de transferencia de aprendizaje como VGG, *ResNet*, *GoogLeNet* o *Inception*. Por ejemplo, en el estudio de Islam et al. (2023) se utilizaron los modelos VGG-16, VGG-19 y *ResNet-50* previamente entrenados con transferencia de aprendizaje para identificar y clasificar enfermedades en 10,000 imágenes de hojas infectadas y sanas de cultivos como tomates, papas y pimientos provenientes de *PlantVillage* obteniendo 98.98% de precisión con el modelo *ResNet-50*. De forma similar, en la investigación de Agarwal et al. (2019) se propone la arquitectura de VGG16 e Inception V3 para identificar hojas sanas, roya común, tizón y mancha gris en 3,852 imágenes de hojas de maíz del conjunto de datos de *PlantVillage*, obteniendo un 94% de precisión.

Por otra parte, Waheed et al. (2020) propone la red neuronal convolucional optimizada, DenseNet, para detectar la roya común, mancha gris y tizón en 12,332 imágenes de hojas de maíz con un 98.06% de resultado de precisión global. De forma complementaria, en la investigación de Amin et al. (2022) se propone *EfficientNetB0* y *DenseNet121* para identificar mancha gris foliar, roya común, tizón y hojas sanas del maíz en imágenes del conjunto de datos de *PlantVillage* obteniendo 98.56% como resultado de precisión.

En el estudio realizado por Yin et al. (2022) se propone el modelo *DISE-Net* para detectar en 1268 imágenes la enfermedad de la hoja pequeña de maíz, causada por el hongo *Bipolaris maydis* con un resultado de 97.12% de precisión. Mientras que, en la investigación de Zhu & Gao (2024) se utiliza *MC-ShuffleNetV2* como variante optimizada de *ShuffleNetV2* con el módulo de atención *Convolutional Block Attention Module* (CBAM) para detectar tizón y roya común en 2,725 imágenes obtenidas en el norte de Anhui, China con 99.86% de precisión. También, en la investigación de Zhou et al. (2024) se utiliza el modelo *ShuffleNetV2*, optimizado con técnicas como *Max Pooling* y el mecanismo de atención *SimAM* para identificar tizón y mancha en 1,902 imágenes de hojas de maíz capturadas en el campo, las cuales fueron aumentadas a 4,000 mediante técnicas de procesamiento de imágenes obteniendo un 98.40% como resultado de precisión. Además, en el estudio de Chen et al. (2021) se utiliza *Mobile-DANet* para detectar la mancha por *phaeosphaeria*, podredumbre

de la mazorca, mancha gris, roya y mancha ocular del maíz en imágenes de *PlantVillage* y del Instituto de Botánica Subtropical de Xiamen, China con resultados de precisión del 98.50% y de 95.86%, respectivamente.

En la investigación de Wang et al. (2022) se desarrolló el modelo *AlexNet*, para detectar roya común, mancha curvularia lunata, tizón del norte y podredumbre del tallo en 2,725 imágenes de la Academia de Ciencias Agrícolas de Anhui, China, las cuales fueron aumentadas mediante técnicas de procesamiento de imágenes para alcanzar un total de 10,785 imágenes obteniendo una precisión promedio del 99.35%. Mientras que, en el estudio de Ishengoma et al. (2021) se identifican agujeros pequeños y manchas traslucidas provocadas por el gusano cogollero en 11,280 imágenes de campos de maíz de Tanzania, utilizando VGG16, VGG19, *InceptionV3* y *MobileNetV2*, con resultados de precisión del 99.92%, 99.67%, 100% y 100%, respectivamente.

De forma similar, Hu et al. (2020) se desarrolla un modelo que combina aprendizaje por transferencia y técnicas de aumento de datos para identificar mancha gris, roya común, tizón foliar y hojas sanas en 4,354 imágenes de hojas de maíz de *PlantVillage* con 98.02% de precisión. También, Çakmak (2024) se enfoca en la clasificación automática de roya común, mancha gris, tizón foliar y hojas sanas utilizando el modelo *EfficientNet B3*, obteniendo 99.66% de precisión. Por su parte, Leng et al. (2023) abordan la detección automatizada del tizón foliar con imágenes RGB capturadas con cámaras de mano y drones con una arquitectura basada en *YOLOv5*, denominada *CEMLB-YOLO*, la cual se validó con indicadores de eficiencia computacional como número de parámetros y *FLOPs* obteniendo como resultado un mAP@0.5 del 87.5%.

La investigación de Pfordt y Paulus (2025) presenta una revisión sistemática, empleando arquitecturas como *AlexNet*, VGG16, *ResNet* y *GoogLeNet* para identificar la plaga del gusano cogollero en imágenes de PlantVillage las cuales se evaluaron con las métricas estándar de precisión, sensibilidad, especificidad y puntaje F1. Por su parte, Reddy et al. (2021) proponen un sistema automatizado para la detección de roya común, mancha foliar y tizón en 1,239 imágenes capturadas con cámaras de teléfonos móviles en entornos de campos agrícolas. Se utiliza el modelo *Faster* R-CNN con el extractor de características *ResNet*-50, el cual fue validado con las métricas *Intersection over Union* (IoU), *Average Precision* (AP) y *mean Average Precision* (mAP), alcanzando una precisión global del 94.06%.

En el estudio de Lv et al. (2020) se propone una red denominada DMS-*Robust AlexNet*, que integra convolución dilatada, convolución multi-escala, activación *PReLU* y el optimizador *AdaBound*, incluyendo la aplicación de la transformada *wavelet* y un algoritmo *Retinex* para identificar hoja sana, mancha por cercospora, roya común y tizón en 5,470 imágenes capturadas en campo y de *PlantVillage* obteniendo una precisión de 98.62%. Por su parte, Handrizal et al. (2024) proponen desarrollar una red neuronal convolucional personalizada y la arquitectura previamente entrenada VGG16, implementando el dropout para prevenir el sobreajuste y el optimizador Adam que mejora el rendimiento del sistema para identificar tizón, roya común y mancha foliar en 3,918 imágenes digitales con 95% de precisión. De forma similar, Craze et al. (2022) abordan la identificación automática de mancha gris, tizón foliar, roya común y mancha phaeosphaeria en imágenes capturadas en campos de maíz, utilizando una red VGG16 modificada, la cual obtuvo una precisión del 73.4% mientras que el modelo entrenado con imágenes de *PlantVillage* obtuvieron un 94.1 % de precisión. Finalmente, Zhang et al. (2018) proponen los modelos *GoogLeNet* y *Cifar10*, con las funciones de activación *ReLU*, capas de regularización *Dropout* y combinaciones específicas de pooling para identificar mancha curvularia, mosaico enano, mancha gris, tizón, mancha parda, mancha redonda, roya y tizón en 3,060 imágenes obtenidas de *PlantVillage* y *Google* con resultados del 98.90% y de 98.8% de precisión respectivamente

Enfoque hibrido

En el trabajo de Qian et al. (2022) se propone un modelo de red neuronal convolucional con un mecanismo de atención y transformación basado en el modelo ViT (*Vision Transformer*) y CCT (*Cascaded Convolutional Transformer*) para eliminar ruido de fondo en las imágenes e identificar tizón foliar, roya común y mancha gris en 7,701 imágenes de *PlantVillage* y fotografías tomadas con teléfonos móviles en campos de maíz, obteniendo un 98.70% de precisión. Por otra parte, Yu et al. (2021) proponen utilizar técnicas de agrupamiento *K-Means* y el modelo VGG-19 para detectar roya y mancha foliar en 900 imágenes del *Crop Disease Recognition of the 2018 Artificial Intelligence Challenger Competition*, obteniendo 93% de precisión. En paralelo, Wahyuningrum et al. (2023) comparan la precisión de un modelo de red neuronal convolucional tradicional con el modelo mejorado *Loss-Fused Convolutional Neural Network* (LF-CNN) para la clasificación de hojas sanas, roya común, tizón y mancha gris con imágenes de *Mendeley Data*, con precisión de 94.15%. Finalmente, Masood et al. (2023) proponen el modelo *Faster*-RCNN llamado *MaizeNet* que emplea los módulos *ResNet*-50 y *Convolutional Block Attention* (CBA) para detectar mancha gris y tizón foliar

en imágenes del *Corn Disease and Severity* del Centro de Agronomía para Investigación y Educación de la Universidad de Purdue con resultados de precisión del 97.91 %.

En la investigación de Santhi et al. (2024) se propone el modelo VGG 16 con clasificadores de *K-nearest neighbors* (KNN) y redes neuronales artificiales (ANN) para detectar las plagas tizón, roya común y mancha gris en imágenes de *PlantVillage* con resultados de precisión del 95.25 %. Por su parte, Xu et al. (2023) proponen un modelo que agrega el mecanismo de atención ECA (*Environment-Cognition-Action*) y el optimizador *Adam* a la red neuronal convolucional *ResNet50* para detectar tizón foliar, mancha gris, roya común, barrenador del tallo y gusano cogollero en 2,650 imágenes de la Universidad Agrícola del Noreste de China. Además, Ji et al. (2024) plantean el modelo *ICS-ResNet*, el cual incorpora los módulos de atención espacial (ISA) y atención de canal (ICA) a la *ResNet-50* para identificar mancha ceciliana, roya común, tizón foliar y virus de la línea en 8,280 imágenes de *PlantVillage*, obteniendo 98.87% de precisión. Finalmente, Kundu et al. (2022) utilizan el algoritmo de agrupamiento *K-Means* y un modelo personalizado llamado *MaizeNet* para detectar tizón foliar y roya común en 2,996 imágenes del *Indian Council of Agricultural Research* (ICAR) el cual, presentó una precisión de 98.50%.

Método de investigación

Debido a que el diagnóstico tradicional realizado por personas para identificar enfermedades y plagas en cultivos de maíz es subjetivo y complejo por las condiciones propias del entorno agrícola, se propone elaborar una plataforma web con un modelo de aprendizaje profundo para identificar plagas en imágenes de hojas de maíz mediante la siguiente metodología de solución:

Etapa 1. Realizar la captura de imágenes con dron y celular

En la primera fase, se realiza una colección de imágenes de hojas de maíz sanas y con plaga de diversos cultivos capturadas con dron y celular.

Etapa 2. Crear conjunto de imágenes capturadas con dron y celular

En la segunda fase, se elabora una base de datos con las imágenes obtenidas en la etapa 1 para clasificar imágenes de hojas maíz sanas o con plaga.

Etapa 3. Estudiar y elaborar una arquitectura de aprendizaje profundo

En la tercera fase, se estudian características y funcionamiento de diversos modelos de aprendizaje profundo para elaborar una con la cual identificar plagas en imágenes de hojas de maíz.

Etapa 4. Entrenar el modelo de Aprendizaje Profundo

En la cuarta fase, se entrena el modelo de aprendizaje profundo elaborado con las imágenes recopiladas y clasificadas para que aprenda a reconocer patrones y realizar predicciones.

Etapa 5. Validar o evaluar el modelo de Aprendizaje Profundo

En la quinta fase, se divide el conjunto de imágenes en entrenamiento, validación y prueba para evaluar el modelo con métricas de clasificación.

Etapa 6. Visualizar y analizar imágenes en una plataforma web.

En la sexta fase se desarrolla una plataforma web en la cual se adjuntan imágenes nuevas para que el modelo de aprendizaje profundo identifique si son de hojas de maíz sanas o con plaga.

Conclusiones

En esta investigación se realiza una búsqueda de trabajos relacionados sobre el uso de algoritmos de inteligencia artificial para detectar enfermedades y plagas en imágenes de hojas de maíz capturadas con dron y celular con el objetivo de proporcionar a los agricultores un recurso informático que les facilite el monitoreo y toma de decisiones sin necesidad de desplazamientos físicos constantes al campo.

La implementación de modelos de aprendizaje profundo, además de apoyar a las medidas de control, promueve la transición hacia la agricultura orgánica al reducir la dependencia de químicos. Por lo tanto, si se mejora el rendimiento de los cultivos sin dañar el ambiente, esta tecnología emergente, contribuye directamente a los Objetivos de Desarrollo Sostenible 2 y 12.

Referencias

Agarwal, M., Bohat, V. K., Ansari, M. D., Sinha, A., Gupta, S. K., & Garg, D. (2019). A Convolution Neural Network based approach to detect the disease in Corn Crop. *Proceedings of the 2019 IEEE 9th International Conference on Advanced Computing*, *IACC 2019*, 176–181. https://doi.org/10.1109/IACC48062.2019.8971602

Amin, H., Darwish, A., Hassanien, A. E., & Soliman, M. (2022). End-to-End Deep Learning Model for Corn Leaf Disease Classification. *IEEE Access*, 10, 31103–31115. https://doi.org/10.1109/ACCESS.2022.3159678

Aravind, K. R., Raja, P., Mukesh, K. V., Aniirudh, R., Ashiwin, R., & Szczepanski, C. (2018). Disease classification in maize crop using bag of features and multiclass support vector machine. *Proceedings of the 2nd International Conference on Inventive Systems and Control*, *ICISC 2018*, Icisc, 1191–1196. https://doi.org/10.1109/ICISC.2018.8398993

Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering*, 172, 84–91. https://doi.org/10.1016/j.biosystemseng.2018.05.013

Brewbaker, J. L., & Hawaii, T. (2003). *Corn Production in the Tropics*. www.ctahr.hawaii.edu

Çakmak, M. (2024). Automatic Maize Leaf Disease Recognition Using Deep Learning. *Sakarya University Journal of Computer and Information Sciences*, *7*(1), 61–76. https://doi.org/10.35377/saucis...1418505

Chen, J., Wang, W., Zhang, D., Zeb, A., & Nanehkaran, Y. A. (2021). Attention embedded lightweight network for maize disease recognition. *Plant Pathology*, *70*(3), 630–642. https://doi.org/10.1111/ppa.13322

Craze, H. A., Pillay, N., Joubert, F., & Berger, D. K. (2022). Deep Learning Diagnostics of Gray Leaf Spot in Maize under Mixed Disease Field Conditions. *Plants*, *11*(15), 1–17. https://doi.org/10.3390/plants11151942

Goodfellow, I., Bengio, Y., & Aaron Courville (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org

Handrizal, Manik, F. Y., & Sianturi, V. J. (2024). Classification of Corn Leaf Diseases Using Cnn: a Deep Learning Approach. *Journal of Theoretical and Applied Information Technology*, *102*(24), 8937–8945.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 770–778. https://doi.org/10.1109/CVPR.2016.90

Hendrycks, D. (2024). *Introduction to AI safety, ethics, and society*. CRC Press.

Hu, R., Zhang, S., Wang, P., Xu, G., Wang, D., & Qian, Y. (2020). The identification of corn leaf diseases based on transfer learning and data augmentation. *ACM International Conference Proceeding Series*, 58–65. https://doi.org/10.1145/3403746.3403905

Huck, W. (2023). Transforming our world: the 2030 Agenda for Sustainable Development. *Sustainable Development Goals*. https://doi.org/10.5040/9781509934058.0025

Idress, K. A. D., Gadalla, O. A. A., Öztekin, Y. B., & Baitu, G. P. (2024). Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing. *Tarim Bilimleri Dergisi*, *30*(3), 464–476. https://doi.org/10.15832/ankutbd.1288298

Ishengoma, F. S., Rai, I. A., & Said, R. N. (2021). Identification of maize leaves infected by fall armyworms using UAV-based imagery and convolutional neural networks. *Computers and Electronics in Agriculture*, *184*(October 2020), 106124. https://doi.org/10.1016/j.compag.2021.106124

Islam, M. M., Adil, M. A. A., Talukder, M. A., Ahamed, M. K. U., Uddin, M. A., Hasan, M. K., Sharmin, S., Rahman, M. M., & Debnath, S. K. (2023). DeepCrop: Deep learning-based crop disease prediction with web application. *Journal of Agriculture and Food Research*, *14*(July). https://doi.org/10.1016/j.jafr.2023.100764

Ji, Z., Bao, S., Chen, M., & Wei, L. (2024). ICS-ResNet: A Lightweight Network for Maize Leaf Disease Classification. *Agronomy*, *14*(7). https://doi.org/10.3390/agronomy14071587

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*(January), 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Krishna, S. T., & Kalluri, H. K. (2019). Deep learning and transfer learning approaches for image classification. *International Journal of Recent Technology and Engineering*, *7*(5), 427–432.

Kundu, N., Rani, G., Dhaka, V. S., Gupta, K., Nayaka, S. C., Vocaturo, E., & Zumpano, E. (2022). Disease detection, severity prediction, and crop loss estimation in MaizeCrop using deep learning. *Artificial Intelligence in Agriculture*, 6, 276–291. https://doi.org/10.1016/j.aiia.2022.11.002

L. Paliwal, R. (2001a). El maíz en los trópicos: mejoramiento y producción. En *Food and Agriculture Organization of the United Nations*. http://www.fao.org/docrep/003/x7650s/x7650s00.htm

L. Paliwal, R. (2001b). El Maíz en los trópicos: mejoramiento y producción. *Fao*, 392. http://www.fao.org/docrep/003/x7650s/x7650s00.htm

Leng, S., Musha, Y., Yang, Y., & Feng, G. (2023). CEMLB-YOLO: Efficient Detection Model of Maize Leaf Blight in Complex Field Environments. *Applied Sciences* (Switzerland), *13*(16). https://doi.org/10.3390/app13169285

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors* (Switzerland), *18*(8), 1–29. https://doi.org/10.3390/s18082674

Lv, M., Zhou, G., He, M., Chen, A., Zhang, W., & Hu, Y. (2020). Maize Leaf Disease Identification Based on Feature Enhancement and DMS-Robust Alexnet. *IEEE Access*, 8, 57952–57966. https://doi.org/10.1109/ACCESS.2020.2982443

Masood, M., Nawaz, M., Nazir, T., Javed, A., Alkanhel, R., Elmannai, H., Dhahbi, S., & Bourouis, S. (2023). MaizeNet: A Deep Learning Approach for Effective Recognition of Maize Plant Leaf Diseases. *IEEE Access*, 11(May), 52862–52876. https://doi.org/10.1109/ACCESS.2023.3280260

Mejía, D. (2003). Maize: Post-Harvest Operation (F. (Technical) AGST/FAO: Danilo Mejía, PhD (ed.).

Pfordt, A., & Paulus, S. (2025). A review on detection and differentiation of maize diseases and pests by imaging sensors. *Journal of Plant Diseases and Protection*, *132*(1). https://doi.org/10.1007/s41348-024-01019-4

Qian, X., Zhang, C., Chen, L., & Li, K. (2022). Deep Learning-Based Identification of Maize Leaf Diseases Is Improved by an Attention Mechanism: Self-Attention. *Frontiers in Plant Science*, *13*(April), 1–15. https://doi.org/10.3389/fpls.2022.864486

Reddy, D. B. R. S., Madhavi, D. G. B., Lakshmi, C. H. S., Nagendra, D. K. V., & Sridevi, D. R. (2021). Detection of Disease in Maize Plant Using Deep Learning. *Alinteri Journal of Agriculture Sciences*, *36*(2), 82–88. https://doi.org/10.47059/alinteri/v36i2/ajas21118

Resti, Y., Irsan, C., Putri, M. T., Yani, I., Anshori, & Suprihatin, B. (2022). Identification of Corn Plant Diseases and Pests Based on Digital Images using Multinomial Naïve Bayes and K-Nearest Neighbor. *Science and Technology Indonesia*, *7*(1), 29–35. https://doi.org/10.26554/sti.2022.7.1.29-35

Rouhiainen, L. (2020). Artificial Intelligence: 101 Things You Should Know Today About Our Future. https://doi.org/10.1128/AAC.03728-14

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence A Modern Approach*. Pearson.

Santhi, S., Murugan, M., Srinivasan, T., & Shanthi, K. G. (2024). Corn Leaf Disease Detection using Deep Learning Techniques. *2nd International Conference on Sustainable Computing and Smart Systems, ICSCSS 2024–Proceedings*, *ICSCSS*, 1536–1540. https://doi.org/10.1109/ICSCSS60660.2024.10625229

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution*, *3*(3), 430–439. https://doi.org/10.1038/s41559-018-0793-y

Unal, Z. (2020). Smart Farming Becomes even Smarter with Deep Learning–A Bibliographical Analysis. *IEEE Access*, 8, 105587–105609. https://doi.org/10.1109/ACCESS.2020.3000175

Waheed, A., Goyal, M., Gupta, D., Khanna, A., Hassanien, A. E., & Pandey, H. M. (2020). An optimized dense convolutional neural network model for disease recognition and classification in corn leaf. *Computers and Electronics in Agriculture*, *175*(January), 105456. https://doi.org/10.1016/j.compag.2020.105456

Wahyuningrum, R. T., Kusumaningsih, A., & Yousi, D. M. (2023). Classification of Corn Leaf Diseases using Loss-Fused Convolutional Neural Network. *Proceedings of 2023 International Conference on Information Management and Technology*, *ICIM-Tech 2023*, 696–701. https://doi.org/10.1109/ICIMTech59029.2023.10277763

Wang, Y., Jin Tao, & Gao, H. (2022). Corn Disease Recognition Based on Attention Mechanism Network. *Axioms*, *11*(480).

Xu, W., Li, W., Wang, L., & Pompelli, M. F. (2023). Enhancing Corn Pest and Disease Recognition through Deep Learning: A Comprehensive Analysis. *Agronomy*, *13*(9). https://doi.org/10.3390/agronomy13092242

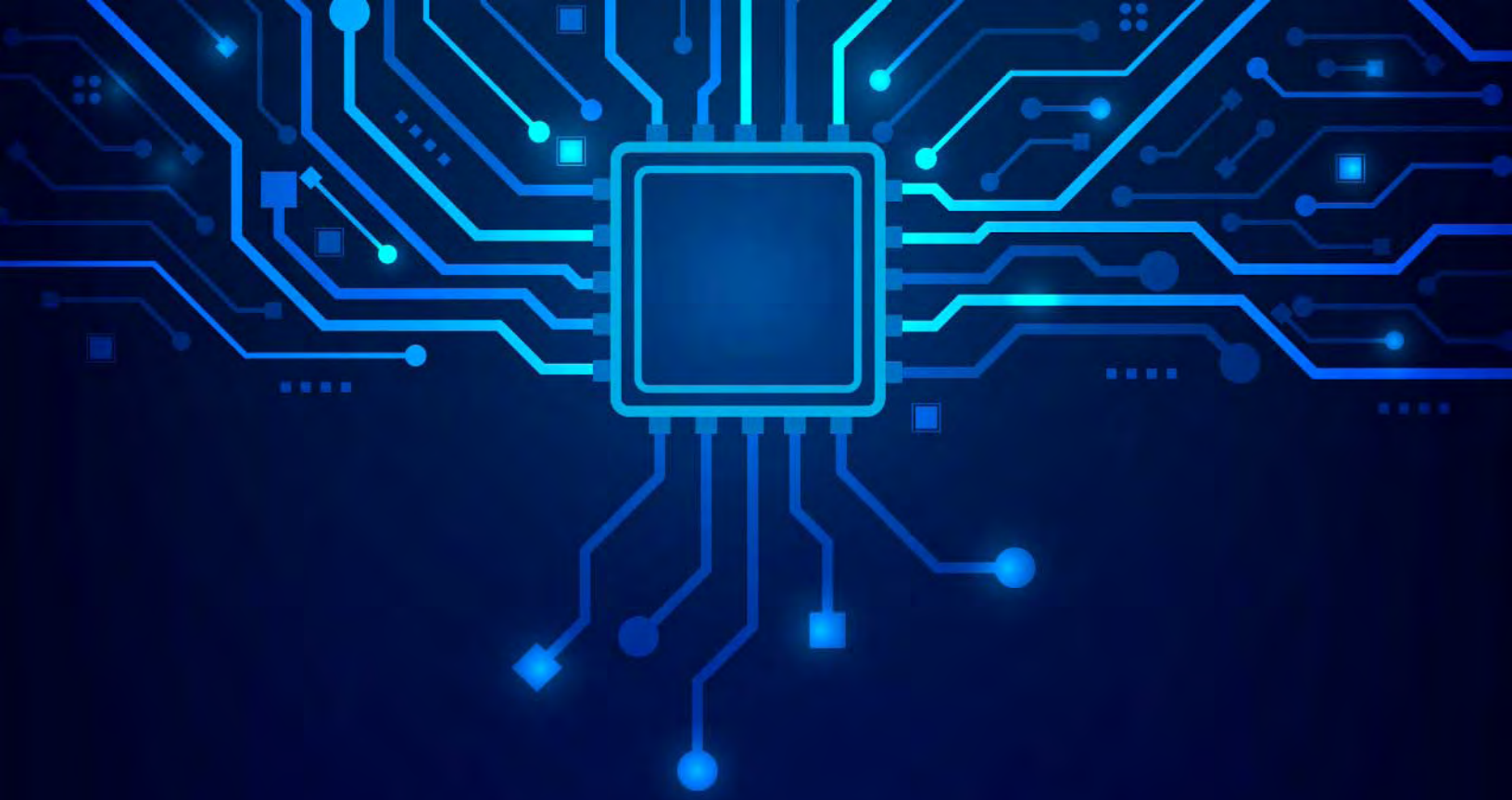
Yin, C., Zeng, T., Zhang, H., Fu, W., Wang, L., & Yao, S. (2022). Maize Small Leaf Spot Classification Based on Improved Deep Convolutional Neural Networks with a Multi-Scale Attention Mechanism. *Agronomy*, *12*(4). https://doi.org/10.3390/agronomy12040906

Yu, H., Liu, J., Chen, C., Heidari, A. A., Zhang, Q., Chen, H., Mafarja, M., & Turabieh, H. (2021). Corn Leaf Diseases Diagnosis Based on K-Means Clustering and Deep Learning. *IEEE Access*, 9, 143824–143835. https://doi.org/10.1109/ACCESS.2021.3120379

Zhang, X., Qiao, Y., Meng, F., Fan, C., & Zhang, M. (2018). Identification of maize leaf diseases using improved deep convolutional neural networks. *IEEE Access*, 6, 30370–30377. https://doi.org/10.1109/ACCESS.2018.2844405

Zhou, H., Su, Y., Chen, J., Li, J., Ma, L., Liu, X., Lu, S., & Wu, Q. (2024). Maize Leaf Disease Recognition Based on Improved Convolutional Neural Network ShuffleNetV2. *Plants*, *13*(12), 1–18. https://doi.org/10.3390/plants13121621

Zhu, S., & Gao, H. (2024). MC-ShuffleNetV2: A lightweight model for maize disease recognition. *Egyptian Informatics Journal*, *27*(January), 100503. https://doi.org/10.1016/j.eij.2024.100503



08.

Triple convergencia en la era de la transformación digital del turismo: ciencia de datos, inteligencia analítica y gestión de destinos

Manuel Ramón González Herrera
Universidad Autónoma de Ciudad Juárez, México
ORCID: 0000-0002-2104-4702

Carolina Medina García
Universidad Autónoma de Ciudad Juárez, México

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.08

08.

Triple convergencia en la era de la transformación digital del turismo: ciencia de datos, inteligencia analítica y gestión de destinos

Introducción

El turismo actual está experimentando transformaciones estructurales radicales como consecuencia de la Cuarta Revolución Industrial, que ha promovido entornos cada vez más inteligentes y automatizados, desvaneciendo así los límites entre lo físico, lo digital y lo virtual (Schwab, 2018). Este influjo ha conducido a la integración de tecnologías inteligentes en la Gestión de Destinos Turísticos (GDT), entre las cuales sobresalen el Internet de las Cosas (IoT), el *Big Data* (BD), la inteligencia artificial (IA) y los Sistemas Automatizados de Gestión. Derivado de esta evolución, ha surgido el concepto de Turismo 4.0, que, según Korže (2019), se entiende como la implementación de los principios y tecnologías de la Cuarta Revolución Industrial en el turismo, con el propósito de crear Destinos Turísticos Inteligentes (DTI), personalizar las experiencias de los turistas y optimizar la gestión de los recursos turísticos mediante la integración entre el mundo físico y el digital.

Consecuente con los nuevos paradigmas, se reconoce que la fusión entre la Industria 4.0 y el desarrollo turístico requiere ecosistemas digitales integrados y nuevos modelos estratégicos de gestión (Abdurakhmanova et al., 2022). Esta transición supone no solo las innovaciones tecnológicas, sino, además, cambios en los modelos de gobernanza turística y la planificación, ordenamiento y gestión del espacio turístico desde una perspectiva integrada, interconectada y digitalizada (Zupan, 2019), lo cual puede contribuir de forma sustancial al fortalecimiento de la sostenibilidad, la optimización del uso de los recursos turísticos y la minimización de los impactos ambientales que se producen a nivel del destino turístico (Bratić et al., 2025).

En este escenario, se identifica como contradicción principal el avance en la capacidad para generar, analizar y aplicar los datos masivos a la GDT, frente a la persistente brecha para su aplicación holística, equitativa, estratégica y ética, toda vez que la literatura reporta su implementación de manera fragmentada, con sesgos algorítmicos, desafíos éticos y una brecha digital en el acceso a las tecnologías. Esta disyuntiva entre el alto potencial tecnológico y la práctica de la gobernanza turística permite reconocer la necesidad de superar la etapa de análisis aislados de las aplicaciones para comprender la convergencia óptima entre estos ámbitos de gestión, lo que plantea la interrogante: ¿Cómo deben converger la ciencia de datos, la IA y la GDT y cuáles son las principales tendencias, aplicaciones y dilemas que reporta la literatura científica? En tanto, el objetivo general es analizar la convergencia entre la ciencia de datos, la inteligencia analítica y la GDT, a partir de la producción científica publicada en *Scopus* durante el periodo 2024-2026.

Los resultados del capítulo se estructuran en dos secciones que permiten comprender de manera integral la intersección entre la tecnología y la GDT. La sección de resultados derivados del análisis de la literatura aborda los fundamentos conceptuales y técnicos de esta relación, desglosándose en tres ejes temáticos: contextualización sobre el turismo en la Era Digital para luego profundizar en la Ciencia de Datos y GDT, destacando la importancia de los datos como activo estratégico para la competitividad y la sostenibilidad, así como la ética en su manejo; finalmente, se explora la IA analítica en la GDT, examinando sus aplicaciones estratégicas y transversales, junto con las técnicas analíticas avanzadas que posibilitan una gestión inteligente de los destinos. Posteriormente, la sección de discusión presenta los hallazgos en un plano más reflexivo y prospectivo, donde se examina la convergencia entre la ciencia de datos, la inteligencia analítica y la GDT, y se identifican las principales tendencias metodológicas, aplicaciones estratégicas y dilemas éticos emergentes que configuran el futuro de este campo de estudio.

Método de investigación

La construcción metodológica de este estudio se basó en una revisión integradora de la literatura, orientada a la búsqueda, el análisis crítico y la reinterpretación de hallazgos previos, lo que permitió identificar tendencias metodológicas, aplicaciones estratégicas y desafíos éticos a considerar. El proceso de búsqueda bibliográfica se realizó de manera intencionada y estratégica, utilizando la base de datos académica *Scopus*, con el propósito de acceder a fuentes revisadas e indexadas que permitieran estudiar la producción científica

en la frontera del conocimiento en IA y en la ciencia de datos aplicada al turismo durante el período comprendido entre 2024 y 2026.

Se encontró un total de 1897 documentos, que fueron procesados. El análisis se apoyó en el gestor de referencias bibliográficas *Zotero*, lo que permitió extraer y categorizar la información en tres dimensiones: las tendencias metodológicas, las aplicaciones a la GDT y los dilemas éticos. La síntesis crítico-comparativa se basó en la búsqueda de patrones recurrentes y de conflictos conceptuales. Este enfoque hizo posible construir el argumento que establece la convergencia entre la IA y la ciencia de datos en la GDT, como un paradigma emergente que busca minimizar los conflictos entre la hiperpersonalización tecnológica y la experiencia humana genuina.

Resultados

El turismo en la era digital

El escenario turístico actual está caracterizado por una contradicción entre el impacto del turismo cuando sobrepasa la capacidad de carga de los destinos turísticos y, por tanto, afecta a las comunidades que residen en esos espacios, y el reconocimiento creciente de la necesidad de transitar hacia un escenario de turismo sostenible/regenerativo e inteligente que conduzca a la minimización de los impactos sociales y ambientales (Muktiarni et al., 2026). Esta problemática, unida a la demanda de experiencias digitales, personalización e inmediatez en los viajes turísticos, ha llevado a la utilización de plataformas digitales, como las agencias de viajes en línea (OTAs), que permiten mejorar las operaciones turísticas mediante métricas de rendimiento relativas a aspectos clave como la ocupación turística o la satisfacción de los visitantes.

Por ende, el proceso de toma de decisiones relativas al turismo se enmarca en una nueva era caracterizada por una evidente automatización, la cual ha aportado valiosas herramientas para la GDT ya que proporciona numerosas fuentes de datos transaccionales, de movilidad, de sensores de IoT, de redes sociales y reseñas en plataformas como *Tripadvisor*, así como estadísticas de uso público que facilitan la disponibilidad de información turística relacionada con los datos climatológicos, de eventos o de la flora y la fauna, los cuales son ahora de mayor acceso para los gestores e interesados. La disponibilidad de estas fuentes de datos permite diseñar estrategias para mitigar los problemas relacionados con el *overtourism*, a la vez que crear experiencias inmersivas personalizadas y promover

emprendimientos turísticos resilientes alineados con los principios de la sostenibilidad y del turismo regenerativo (Maiurro et al., 2026).

Esta transformación en la que estamos inmersos permite mejorar la GDT, tanto operativa como estratégica, y, al mismo tiempo, promover estrategias que conduzcan a la consolidación de un desarrollo turístico más sostenible e inteligente. Tal como afirman Muktiarni et al. (2026), el uso de estas herramientas digitales puede contribuir a fortalecer la educación vocacional turística y, en consecuencia, promover el pensamiento turístico consciente mediante la formación de los futuros profesionales basada en enfoques digitales novedosos aplicados a la gestión turística sostenible. Al mismo tiempo, Maiurro et al. (2026) resaltan la contribución de las tecnologías digitales a facilitar la creatividad e innovación social en el sector turístico, lo que trasciende la frontera de los beneficios económicos hacia las dimensiones ambientales, sociales y cognitivas del desarrollo turístico, en respuesta a los Objetivos de Desarrollo Sostenible (Amar, 2024).

No obstante, la implementación de estas buenas voluntades a favor de la transformación digital de la GDT se encuentra frente a obstáculos que limitan su alcance. Según estudios realizados en diversos contextos, los procesos de transformación digital en el turismo requieren un encauzamiento progresivo y consciente de la alfabetización digital y de la adopción de nuevos enfoques tecnológicos, lo cual, según Aytuna y Özkaya (2026), se ve frenado por el acceso desigual y las limitadas competencias digitales, lo que genera una importante brecha digital que afecta la gestión turística inteligente. Otros autores, como Suratno et al. (2026), destacan que la competencia digital afecta directamente los niveles de intención de uso y de satisfacción con el manejo de las tecnologías digitales en el turismo, por lo que, si no se les presta una atención responsable, no se podrán alcanzar los objetivos de inclusión y resiliencia en el sistema turístico. Tales evidencias ponen de manifiesto que, para promover la digitalización en la GDT es imprescindible superar las brechas actuales de acceso, de educación y de recursos necesarios para un desempeño competente.

Ciencia de datos y gestión de destinos turísticos

Los datos se han convertido en un activo estratégico clave para la gestión sostenible e inteligente de los espacios turísticos. La literatura reporta que los destinos que buscan, analizan, interpretan y aprovechan los datos como ventaja competitiva toman mejores decisiones y alcanzan un posicionamiento favorable en un mercado turístico cada vez más competitivo (Cheng & Kannan, 2026). Los datos, en su función de activo fundamental para

cualquier ámbito turístico, abarcan desde las huellas digitales registradas en redes sociales y plataformas especializadas (Gao et al., 2026) hasta datos de movilidad en tiempo real (Wei et al., 2026) y mediciones socioambientales predictivas de gran utilidad (Liu et al., 2026).

Es además de gran valor utilitario en esta era digital el análisis de sentimientos avanzados (Srianan et al., 2025) y el mapeo de la imagen de los destinos utilizando datos de las redes sociales (Rhizlane et al., 2026) así como la elaboración de modelos sobre patrones de consumo particularmente para segmentos emergentes (Chen et al., 2026), lo que hace posible una mejor adaptación preventiva de la oferta y la infraestructura con el fin de favorecer una experiencia memorable al visitante.

De igual manera, la gestión estratégica de datos es fundamental para alcanzar la sostenibilidad de los destinos turísticos en distintos horizontes temporales. Los hallazgos de Qu et al. (2026) reconocen que los datos analizados e interpretados en correspondencia con marcos analíticos consistentes ayudan a cuantificar riesgos, crisis e impactos, lo que facilita la adopción de estrategias de recuperación más resilientes. De igual forma, el uso de métricas relacionadas con los datos obtenidos es muy útil para la planificación de los espacios turísticos, la evaluación de la calidad visual relacionada con la asimilación de cada sitio y la mejor comprensión de las dinámicas que se dan en el espacio turístico, lo que facilita un manejo optimizado de los flujos de visitantes (Liu et al., 2026) y, por tanto, evita los efectos del overtourism. Estos y muchos otros ejemplos sobre el valor de los datos para la GDT permiten promover una gobernanza turística más informada y objetiva en la toma de decisiones, con el fin de contrarrestar los efectos negativos de la saturación del destino.

El creciente valor de este activo ha llevado a posicionar la ciencia de datos como un área del saber de gran valor para la GDT, a pesar de que sus bases, relacionadas con los análisis numéricos y las estadísticas, son mucho más anteriores. Por tal motivo, la ciencia de datos es fundamental para optimizar la GDT con un enfoque ya no tan intuitivo, sino más predictivo y sustentado en evidencias a escala masiva.

En tal sentido, los análisis de datos masivos se obtienen de fuentes muy diversas, como tarjetas inteligentes (Wei et al., 2026), flujos de video (Liu et al., 2026), redes sociales (Cheng & Kannan, 2026), reseñas en plataformas en línea (Gao et al., 2026), entre otras, que permiten modelar el comportamiento de los visitantes. Este enfoque permite reconocer patrones de comportamiento (Chen et al., 2026), identificar y comprender relaciones causa-efecto y establecer recomendaciones de manejo, poniendo de manifiesto el ciclo de la

ciencia de datos desde su etapa descriptiva y de diagnóstico hasta las etapas predictivas y los marcos de acción para la resiliencia (Qu et al., 2026).

La transformación analítica en la era digital está impactando de forma creciente los viajes turísticos, ya que favorece la personalización segmentada y una mejor comprensión en tiempo real de las experiencias de los visitantes utilizando por ejemplo el Aprendizaje Automático (AA) y Procesamiento del Lenguaje Natural (PLN) para el análisis de sentimientos y contenidos generados por los visitantes (Srianan et al., 2025; Gao et al., 2026), así como el uso de datos geoespaciales y temporales para la gestión de flujos turísticos o del impacto ambiental (Wei et al., 2026; Liu et al., 2026).

Como parte de la contribución de la ciencia de datos a la GDT, es posible realizar el Análisis Exploratorio de Datos (EDA) para identificar patrones, anomalías o relaciones entre variables turísticas (Muktiarni et al., 2026), utilizando herramientas como *Python* o *R Studio*, tanto para datos estructurados como para datos no estructurados provenientes de redes sociales, sensores de IoT o reseñas digitales de los visitantes. Tales procesos se sustentan en enfoques metodológicos sólidos y éticos, basados en métricas y KPIs propios de la actividad turística, como la ocupación hotelera o el gasto promedio de los turistas (Maiurro et al., 2026).

Claro está que aún persisten muchas barreras que impiden la plena implementación de la ciencia de datos en la GDT, tales como la brecha digital entre actores, la inequidad tecnológica y de acceso a las herramientas, así como cuestiones de privacidad en el uso de los datos. Esta problemática pone de relieve la necesidad de abordar los desafíos éticos en el tratamiento de datos como parte de la GDT, especialmente aquellos relacionados con la privacidad y los sesgos algorítmicos, evitando tecnologías intrusivas, la discriminación, las inequidades o la falta de transparencia, lo cual han señalado autores como Wang & Lee (2026).

Inteligencia artificial analítica en la gestión de destinos turísticos

La implementación de la IA en la GDT ha evolucionado desde el análisis básico de datos estructurados hasta la gestión inteligente de flujos de información complejos y de naturaleza múltiple. Según Cheng & Kannan (2026) y Yu & Cheng (2025), esta progresión consiste en reconocer enfoques integradores y herramientas de análisis que involucran diferentes tipos de datos —textuales, visuales, geoespaciales o temporales—, con el fin

de prever patrones de comportamiento y mejorar la satisfacción y el bienestar de los turistas. Al respecto, Srianan et al. (2025) reconocen que la capacidad de análisis inteligente de datos está potenciada con técnicas de aprendizaje profundo (*Deep Learning*) y modelos de lenguaje (LM) para temas como el análisis de sentimientos, a lo cual agrega Gao et al. (2026) la evaluación automatizada de la calidad de los datos e información generada por los turistas, a la vez que señala la necesidad de enfrentarse de forma creativa a la sobrecarga de información y a la desinformación.

En particular, la inteligencia artificial analítica es una rama de la IA especializada en analizar grandes volúmenes de datos estructurados y no estructurados para descubrir patrones ocultos, generar información procesable y predecir tendencias futuras, apoyándose en técnicas como el aprendizaje automático, el procesamiento del lenguaje natural y la minería de datos para transformar la información cruda en conocimiento estratégico que facilita la toma de decisiones en entornos empresariales, científicos y sociales. En ámbitos como la planificación estratégica del turismo, la inteligencia analítica se revela como una herramienta fundamental para las decisiones basadas en datos, lo que permite realizar estudios de vulnerabilidad y resiliencia, así como análisis de escenarios y del comportamiento espaciotemporal de los turistas (Chen et al., 2026). Destaca también la capacidad analítica de estas herramientas inteligentes para proyectar mejores modelos de ordenamiento territorial y de las dinámicas del paisaje (Liu et al., 2026), así como para realizar estudios de percepción e imagen del lugar visitado (Rhizlane et al., 2026).

En este orden, la producción científica relacionada con la inteligencia analítica para la GDT pone de relieve que las interacciones entre los humanos y la IA son un factor determinante que deberá considerarse en el escenario turístico actual y futuro, razón por la que la gestión de datos y la inteligencia analítica confluyen en el manejo automatizado de la información y el análisis avanzado como condición para una gobernanza turística que transite desde los enfoques reactivos hacia modelos adaptativos, predictivo-proyectivos y de personalización avanzada. Al mismo tiempo, estudios realizados por Wang & Lee (2026) sobre el turismo oscuro plantean los límites de la autenticidad y de la conexión emocional con el lugar de visita cuando esta se ve mediada por agentes artificiales, lo que evidencia la necesidad de sistematizar estudios que aborden esta problemática.

De tal forma, la inteligencia analítica se convierte en un pilar rector para la GDT, ya que permite transformar datos heterogéneos en información y conocimiento estratégicos, lo que promueve el avance desde la toma de decisiones empíricas hacia las predictivas y

las simuladas. Conectados con esta idea, autores como Cheng y Kannan (2026) proponen modelos que integran texto, imágenes y series temporales para predecir los flujos turísticos, evaluar estrategias de resiliencia o comprender fenómenos turísticos emergentes en los destinos. Todos estos avances reportados en la literatura favorecen una mejor gestión de los flujos turísticos, la calidad de la experiencia turística y la determinación de la capacidad de carga, lo que repercute en una gobernanza basada en evidencia científica y en la agregación de valor a la GDT.

Por ello, en el escenario digital en el que nos encontramos, se emplean numerosas técnicas analíticas avanzadas para la GDT, que integran métodos de análisis predictivo y de aprendizaje automático (*Machine Learning*) y favorecen la planificación y el funcionamiento de las operaciones turísticas. La literatura científica presenta diversos modelos de regresión útiles para realizar predicciones, así como algoritmos de clasificación, con múltiples aplicaciones, como la gestión de demandas, la predicción de cancelaciones o la caracterización de perfiles de visitantes (Muktiarni et al., 2026).

Asimismo, se reporta su validez para el análisis de tendencias en la distribución estacional de las demandas o riesgos y crisis, mediante técnicas como las series de tiempo, que emplean métodos estadísticos para analizar datos históricos, tales como las llegadas de turistas, lo que favorece el reconocimiento de patrones, tendencias y estacionalidades (Song et al., 2019). La asistencia computacional facilita a la GDT metodologías basadas en el *Data Mining*, el Procesamiento de Lenguaje Natural (NLP) y las APIs (Interfaces de Programación de Aplicaciones) de diferentes plataformas para el procesamiento de datos como reseñas, precios, sentimientos de los visitantes o tendencias del mercado (Maiurro et al., 2026).

El valor de uso real de los análisis avanzados de datos por sí solo no logra los resultados esperados, es necesario hacer una visualización efectiva de los datos y storytelling estratégicos que puedan comunicar a las partes interesadas tales hallazgos. Para este propósito se utilizan herramientas modernas como *Tableau*, librerías de *Python* o *Power BI*, con el fin de generar dashboards interactivos al alcance de los diferentes gestores, las empresas turísticas o los hoteleros, para los cuales se ofrece la posibilidad de monitorización en tiempo real mediante los KPIs más críticos (Kirk, 2016).

Los paneles disponibles para la gestión permiten mostrar desde métricas hasta la conversión de la información compleja en narrativas más comprensibles para las autoridades

locales, los inversionistas y los propios visitantes, logrando así una comunicación visual inequívoca como parte de la gestión inteligente y resiliente del destino (García, 2025).

Discusión

Convergencia entre ciencia de datos, inteligencia analítica y gestión de destinos turísticos

El estudio de la producción científica sobre este tema pone de manifiesto la convergencia entre la ciencia de datos, la inteligencia artificial analítica y la GDT, orientación que ha evolucionado desde el uso de aplicaciones dispersas hasta la integración de la gestión digital en el paradigma del turismo inteligente. En este punto, los modelos publicados por Cheng & Kannan (2026) para la predicción de tendencias y por Qu et al. (2026) sobre la resiliencia revelan cómo los algoritmos de machine learning integran, al mismo tiempo, técnicas estadísticas y econométricas más tradicionales en función de la GDT.

Sobre la base de esta triple convergencia, es factible estudiar comportamientos, como demuestran Wei et al. (2026): diagnosticar problemáticas latentes de los destinos o simular escenarios espaciotemporales, por lo que los datos se transforman en conocimiento tangible para la gestión estratégica de los destinos, la minimización de riesgos e impactos, así como la adecuación de la oferta turística a los perfiles específicos de demanda. Esta nueva dinámica está, por tanto, redefiniendo el objeto de estudio y las escalas de análisis en la GDT.

Muestra de ello es que ahora la ciencia de datos y la IA hacen posible el monitoreo e intervención en tiempo real de variables y dimensiones de estudio que antes resultaban poco accesibles para el estudio cuantitativo, lo que reduce los niveles de subjetividad e incertidumbre. Al respecto, Liu et al. (2026) sostienen que la visión por computadora y el aprendizaje profundo son herramientas efectivas para medir la calidad visual de los paisajes urbanos; de igual forma, Gao et al. (2026) defienden en su estudio el empleo de modelos de lenguaje grande (LLMs) para la evaluación automatizada de datos masivos relacionados con las reseñas que dejan los turistas en diversas plataformas. Coincidente con este análisis, Wang & Lee (2026) aportan conocimiento relevante sobre las dimensiones cognitivo-emocionales derivadas de las interacciones humano-IA. Como colofón de esta triple convergencia, Rhizlane et al. (2026) reconocen que el mapeo de datos masivos y el análisis de sentimientos contribuyen positivamente a la GDT.

Tendencias metodológicas, aplicaciones estratégicas y dilemas éticos emergentes

Desde una perspectiva metodológica, se aprecia que el horizonte actual de la producción científica en el ámbito turístico está marcado por un vuelco decisivo desde las fuentes de datos tradicionales y estructuradas, como los datos transaccionales, hacia la complejidad de los datos multimodales, que abarcan diversos tipos de datos. Esta incipiente tendencia está dando paso a la integración de datos textuales, imágenes, videos y geoespaciales, lo que permite una comprensión más holística y profunda de los objetos, fenómenos y procesos turísticos. Entre otras contribuciones, el trabajo de Cheng y Kannan (2026) presenta modelos híbridos que integran datos textuales y visuales para hacer predicciones tendenciales, lo que demuestra que el análisis basado en diferentes formatos ofrece una mayor capacidad predictiva para la GDT.

Al mismo tiempo, se aprecia que, en consonancia con la diversificación de los datos potentes, el aprendizaje profundo y los grandes modelos de lenguaje (LLMs) se van consolidando como eje central y orientador de los análisis. De igual forma, técnicas y procedimientos estadísticos convencionales, así como los algoritmos tradicionales de *machine learning*, se están reemplazando por arquitecturas neuronales avanzadas. Autores como Srianan et al. (2025) aportan evidencia de estudios de sentimientos basados en datos turísticos que reflejan alta precisión y manejo de la complejidad mediante una modelación más avanzada, que supera a modelos anteriores de deep learning.

Otra tendencia se relaciona con el énfasis en la capacidad explicativa y en el análisis de las relaciones causales, con el fin de superar la predicción no interpretable y avanzar hacia una mejor comprensión automatizada de los mecanismos subyacentes a los objetos, fenómenos y procesos turísticos. Destaca la contribución de Qu et al. (2026) al combinar la econometría con el machine learning para encontrar explicaciones causales que orienten la formulación de políticas de resiliencia, y la de Gao et al. (2026), quienes defienden el valor de los LLMs para realizar evaluaciones automáticas de datos masivos de reseñas que, de forma manual, serían casi imposibles.

En cuanto a las aplicaciones estratégicas revisadas, la literatura señala que la producción científica más reciente puede traducirse en herramientas valiosas para una GDT proactiva, sostenible e inteligente, contribuyendo así a una mejor gestión de los flujos turísticos y al control preventivo del overtourism. Se revela, por su alto potencial, el valor de los datos de movilidad reportados por GPS, tarjetas inteligentes y sistemas MaaS, que

ofrecen una movilidad más fluida y sostenible mediante la integración del transporte público y privado en una misma aplicación, lo que permite monitorear y predecir en tiempo real situaciones indeseables, tal como han demostrado Wei et al. (2026). El desarrollo de estas capacidades permite implementar herramientas para la alerta temprana ante riesgos y eventos que requieren respuesta inmediata, así como para modelar situaciones como el exceso de cargas turísticas, con el fin de mejorar la experiencia turística y contribuir a una gestión optimizada con un enfoque preventivo y flexible.

En términos de promoción y comunicación turística, se advierte la consolidación del márketing y la administración de la reputación basados en datos digitales frente a los métodos tradicionalmente utilizados, como las encuestas, lo que promueve la integración de técnicas avanzadas de análisis de las percepciones y opiniones en línea disponibles en plataformas sociales, tal como han demostrado Srianan et al. (2025). En este orden, el mapeo semántico defendido por Rhizlane et al. (2026) para identificar atributos que los visitantes perciben como identitarios del lugar permite un mejor diseño de estrategias y acciones de promoción, sobre todo ante opiniones negativas que afectan la imagen del destino. El quehacer académico evidencia una interesante orientación hacia la adaptación exhaustiva de las experiencias turísticas y la planificación sostenible a lo largo del tiempo. Otras evidencias relacionan la GDT con técnicas de clustering y modelos de comportamiento de los visitantes (Chen et al., 2026), así como con herramientas para la evaluación paisajística a partir de datos digitales (Liu et al., 2026) y con el modelado de resiliencia (Qu et al., 2026). Tales avances tecnológicos a favor de la GDT proporcionan a los tomadores de decisiones información rigurosa para proyectar estrategias basadas en datos.

Finalmente, a partir de la revisión de la producción científica, se reconoce que la digitalización turística se enfrenta a tres dilemas éticos principales. Primeramente, debe reconocerse el desequilibrio que afecta los aspectos relacionados con la privacidad y la personalización, el cual debe ajustarse a los derechos establecidos en los marcos institucionales y regulatorios con los que opera el turismo, para alcanzar la transparencia necesaria y el consentimiento informado de todos los participantes. En segundo lugar, perduran sesgos algorítmicos que acentúan las desigualdades sociales, la discriminación basada en el precio, los sistemas de recomendación no inclusivos o el análisis de sentimientos distorsionado, todo lo cual exige sistemas de gestión de datos interpretables y una supervisión periódica. En tercer lugar, se advierte la problemática que caracteriza la autenticidad de la experiencia turística y la necesidad de una gobernanza más incluyente, aspectos reportados por Wang

y Lee (2026) y Aytuna y Özkaya (2026) respecto al impacto de la IA sobre las respuestas emocionales y a la brecha digital que es necesario salvar para alcanzar una transformación digital justa y ética.

Conclusiones

Se concluye que la revisión integradora de la literatura evidencia que la GDT se encuentra en un momento de transformación integral, incitada por la ciencia de datos y la IA, lo que pone de manifiesto la triple convergencia entre estas áreas de conocimiento y constituye un espacio interdisciplinar emergente en el que la integración de datos y modelos predictivos redefine los procesos de planificación y sostenibilidad. De tal forma, se posiciona en el límite del conocimiento para el tratamiento de problemas complejos del turismo mediante el uso de evidencias, lo que permite explorar cómo las tecnologías pueden contribuir a la transformación de las interacciones entre los diferentes actores que operan en el sistema turístico.

Los tres pilares antes mencionados no emergen como ámbitos independientes y fragmentados, sino que se articulan entre sí, lo que potencia una espiral de mejora continua que reconfigura los modelos actuales de gobernanza de los destinos. Tales avances en el conocimiento científico a favor de mejores prácticas de gestión turística permiten realizar desde el monitoreo en tiempo real hasta la personalización de experiencias, pero este progreso para la GDT supone al mismo tiempo una contradicción que debe ser atendida, ya que demanda de actuaciones éticas que tomen en consideración factores sensibles como la privacidad, equidad y autenticidad como principio básico para que la innovación tecnológica se implemente de manera humana, inclusiva y sostenible.

Se comprueba que la investigación turística reciente (2024-2026) permite reconocer una transformación creciente orientada a la propuesta y a la adopción de modelos de predicción impulsados por la inteligencia analítica y los datos multimodales. Estos hallazgos permiten asumir estrategias de gestión de flujos, de demandas y de experiencias inmersivas de una forma nunca vista antes. En función de este reconocimiento es necesario retomar desafíos imperativos para la consolidación de DTI como la gran brecha digital, la carencia de sistemas interpretables, y que los desarrollos tecnológicos progresan a un ritmo exponencial, por lo que se requiere no solo de la propia innovación en tecnologías sino también de alfabetización digital urgente, nuevos marcos de ética para un mejor uso de estas tecnologías, y de una gobernanza participativa que integre a todos los actores que deben comprometerse con el Turismo 4.0.

Finalmente, es necesario admitir que la revisión bibliográfica integradora, presentada como primer acercamiento al tema, presenta limitaciones, entre ellas la falta de exhaustividad de la revisión de la producción científica internacional, conducida a la luz del marco conceptual de la triple convergencia, por lo que se reconocen posibles sesgos en la identificación de fuentes y en el énfasis de los estudios consultados. Se trata de un tema muy reciente y en evolución, en el que, día a día, emergen nuevas propuestas tecnológicas que apenas dan tiempo de abordarlas, por lo que no se llega a profundizar como se debería en propuestas y buenas prácticas. Claro que estas limitaciones no restan valor a los hallazgos presentados, sino que ofrecen, como síntesis reflexiva, una apertura al debate, ya que sistematizan resultados recientes que han permitido identificar patrones emergentes e ideas para el saber actual y la proyección de estudios futuros.

Referencias

Abdurakhmanova, G., Astanakulov, O., Goyipnazarov, S., & Irmatova, A. (2022). Tourism 4.0: Opportunities for applying Industry 4.0 technologies in tourism. En *Proceedings of the 6th International Conference on Future Networks & Distributed Systems* (*ICFNDS 2022*). Association for Computing Machinery. https://doi.org/10.1145/3584202.3584208

Amar, V. (2024). Educación digital en el marco de los Objetivos de Desarrollo Sostenible. Una narrativa. *Revista Lusófona de Educação*, *61*(61). https://doi.org/10.24140/issn.1645-7250.rle61.10

Aytuna, E. N., & Özkaya, A. E. (2026). Digital Transformation of Rural Women's Cooperatives in Turkey: Challenges and Opportunities. *European Public and Social Innovation Review*, 11. https://n9.cl/20ftc0

Bratić, M., Marić Stanković, A., Pavlović, D., Pivac, T., Kovačić, S., Surla, T., Cerović, S., & Zlatanov, S. (2025). New era of tourism: Innovative transformation through Industry 4.0 and sustainability. *Sustainability*, *17*(9), 3841. https://doi.org/10.3390/su17093841

Chen, X., Tang, S., & Liu, C. (2026). Digital-driven spatial organization of special forces tourism: A multi-dimensional coupling analysis. *Applied Geography*, 186, 103857. https://doi.org/10.1016/j.apgeog.2025.103857

Cheng, P., & Kannan, S. A. L. N. (2026). MM-HAL: A Multi-Modal Data Fusion Model for Tourism Trend Prediction Using Twitter and Flickr Data. *Journal of Circuits, Systems and Computers*, *35*(5), 2550429. https://doi.org/10.1142/S0218126625504298

Gao, J., Peng, P., Claramunt, C., & Lu, F. (2026). Quality Assessment of Tourist Generated Contents: A Large Language Model Approach. En *Lecture Notes in Computer Science: Vol. 15946 LNCS* (pp. 31–41). Springer. https://doi.org/10.1007/978-3-032-01723-9_3

García, Y. D. (2025). Inteligencia artificial en la gestión de destinos turísticos. *Visión Académica*, *1*(1), 1–20. https://doi.org/10.70577/ybq6cm11

Kirk, A. (2016). *Data visualisation: A handbook for data-driven design* (2nd ed.). SAGE Publications.

Korže, S. Z. (2019). From Industry 4.0 to Tourism 4.0. *Innovative Issues and Approaches in Social Sciences*, *12*(3). 10.12959/issn.1855-0541.IIASS-2019-no3-art3

Liu, Z., Su, T., & Jiang, H. (2026). Let the landscape dance: Spatial dynamic changes of visual elements in urban park landscapes based on video panoramic segmentation. *Environmental Impact Assessment Review*, 117, 108225. https://doi.org/10.1016/j.eiar.2025.108225

Maiurro, M., Breda, Z., Brandão, F., & Costa, C. (2026). The role of digital technologies in promoting social innovation in tourism Brazilian companies. *European Public and Social Innovation Review*, 11. https://doi.org/10.31637/epsir-2026-1979

Qu, M., Fu, Y., Chen, D., Zhang, Y., & Bao, J. (2026). From impact to action: Enhancing international tourism resilience through counterfactual explanations. *Tourism Management*, 114, 105356. https://doi.org/10.1016/j.tourman.2025.105356

Rhizlane, S., Abyre, A., & Abdelmajid, C. (2026). Mapping research on the influence of social networks on tourist destination image. Trends and future directions (2015-2024). *Multidisciplinary Science Journal*, *8*(5), e2026307. https://doi.org/10.31893/multiscience.2026307

Schwab, K. (2018). The Fourth Industrial Revolution. En *Encyclopædia Britannica Anniversary Edition: 250 Years of Excellence* (*1768–2018*). Encyclopædia Britannica. https://www.britannica.com/event/The-Fourth-Industrial-Revolution-2119734

Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75, 338–362. https://doi.org/10.1016/j.annals.2018.12.001

Srianan, S., Nanthaamornphong, A., & Phucharoen, C. (2025). Advancing tourism sentiment analysis: a comparative evaluation of traditional machine learning, deep learning, and transformer models on imbalanced datasets. *Information Technology and Tourism*, *27*(4), 1011–1045. https://doi.org/10.1007/s40558-025-00336-0

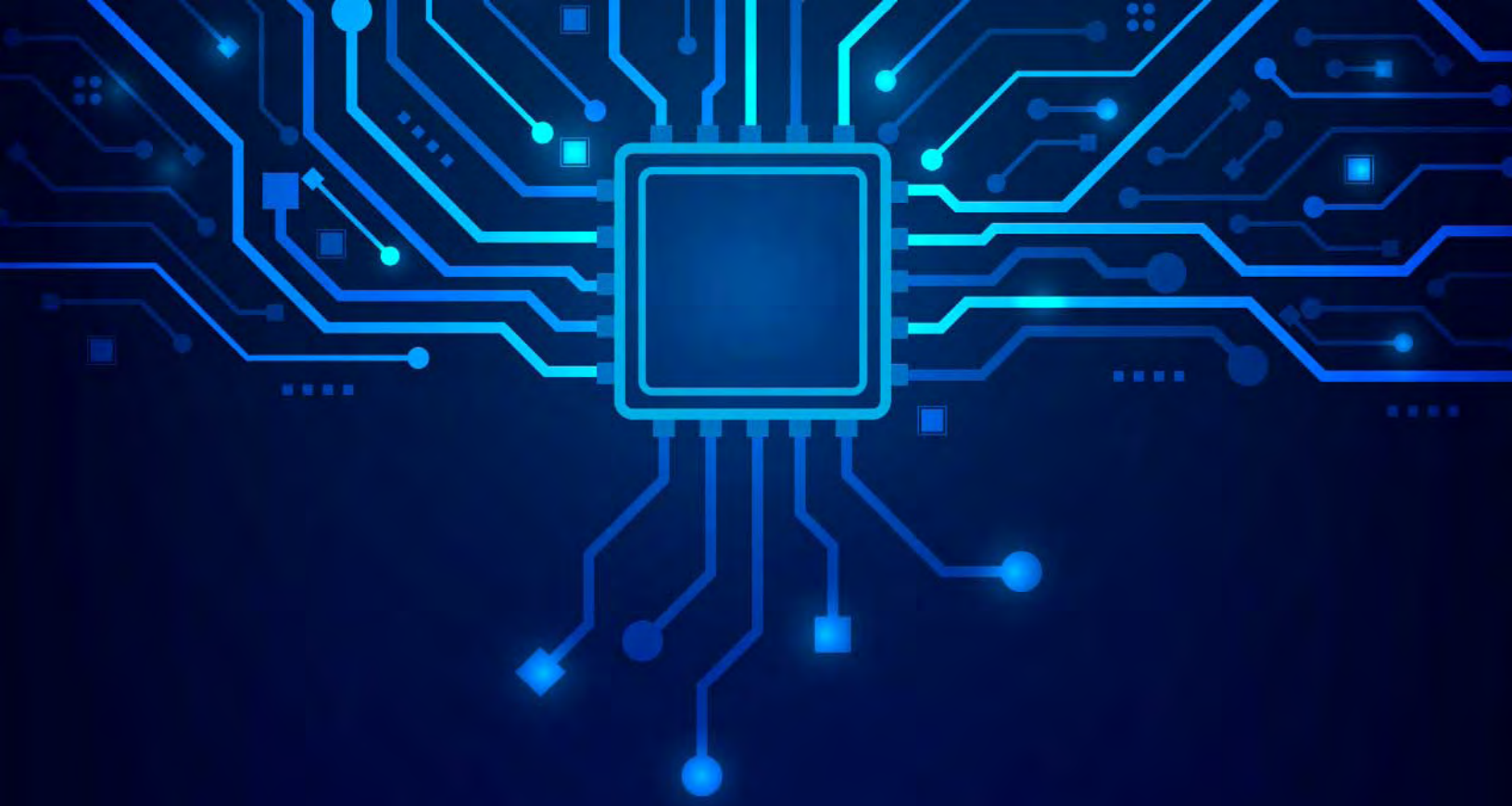
Suratno, T., Johannes, Marzal, J., & Oktavia, A. (2026). Post-adoption behavior analysis: An empirical study of the expectation confirmation model on sharia mobile banking in Indonesia. *Multidisciplinary Science Journal*, *8*(2), e2026135. https://doi.org/10.31893/multiscience.2026135

Wang, H., & Lee, T. J. (2026). Narration in dark tourism: Mediating and moderating effects in tourists' responses to AI versus human voices. *Journal of Hospitality and Tourism Management*, 66, 101367. https://doi.org/10.1016/j.jhtm.2025.101367

Wei, J., Kan, Z., & Kwan, M.-P. (2026). Revealing sequential activity-travel patterns and spatial structures among older and younger adults from smart card data. *Transportation Research Part A: Policy and Practice*, 203, 104749. https://doi.org/10.1016/j.tra.2025.104749

Yu, X., & Cheng, M. (2025). Multimodality in tourism and hospitality: A critical and narrative review. *Tourism Management*, 111, 105245. https://doi.org/10.1016/j.tourman.2025.105245

Zupan, S. (2019). From Industry 4.0 to Tourism 4.0. *Innovative Issues and Approaches in Social Sciences*, *12*(3), 29–52.



09.

Uso de la inteligencia artificial en el coaching empresarial (2024): revisión de literatura

Arturo González Torres
Instituto Tecnológico de Milpa Alta / Tecnológico Nacional de México
ORCID: 0000-0002-3337-7600

Gerardo Quiroz Bojorges
Instituto Tecnológico de Milpa Alta / Tecnológico Nacional de México
ORCID: 0000-0003-0990-1908

Pavel David Ulises Avendaño López
Instituto Tecnológico de Milpa Alta / Tecnológico Nacional de México
ORCID: 0000-0003-1536-1784

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.09

09.

Uso de la inteligencia artificial en el coaching empresarial (2024): revisión de literatura

Introducción

El Planteamiento del problema

La inteligencia artificial (IA) se ha consolidado como un eje central de transformación tecnológica, social y organizacional, al permitir la automatización de tareas y el apoyo a la toma de decisiones en múltiples sectores, incluidos el ámbito laboral y las organizaciones. La transformación en los procesos laborales y en los modelos de gestión organizacionales tradicionales genera incertidumbre en cuanto a la sustitución de funciones humanas, la reconfiguración de competencias laborales y la dependencia de la automatización en los sistemas de toma de decisiones.

En estos casos, uno de los problemas lo constituye la incorporación de la IA en los dilemas tecnológicos, éticos, sociales y humanos, como la pérdida o transformación de empleos y la discriminación algorítmica, la falta de transparencia en la toma de decisiones y la necesidad de dar esos principios y esa responsabilidad.

Objetivo

El objetivo general de la investigación es analizar, desde un enfoque teórico, las implicaciones de la IA y el *coaching* organizacional como estrategia para gestionar los procesos de cambio derivados de la transformación digital.

Justificación

Esta investigación se justifica por la creciente interacción entre la IA y el *coaching* organizacional en los contextos laborales contemporáneos, ya que la incorporación de tecnologías basadas en IA ha transformado los procesos de trabajo y la toma de decisiones,

generando tanto oportunidades de eficiencia e innovación como desafíos relacionados con el empleo, la ética y la gestión del talento humano. En este escenario, el *coaching* emerge como una variable clave para acompañar la adaptación de las personas y los líderes a entornos altamente digitalizados, fortaleciendo competencias, bienestar emocional y liderazgo consciente frente al uso creciente de sistemas automatizados.

Aunque todavía no se entiende del todo cómo las dos variables pueden describirse de forma complementaria para reforzar la efectividad organizacional y minimizar el impacto negativo de la transformación digital, esta investigación, elaborada bajo la metodología Colin (2007), pretende examinar la relación entre la IA y el *coaching* para apoyar la elaboración de un marco conceptual que facilite una gestión organizacional más ética, humanitaria y sostenible, dadas las circunstancias.

Antecedentes

La inteligencia artificial y sus implicaciones en el empleo y la toma de decisiones

La IA se ha consolidado como uno de los principales motores de transformación tecnológica y social del siglo XXI, debido a su capacidad para simular procesos cognitivos humanos y automatizar tareas complejas mediante el análisis avanzado de datos. Gracias al aumento de la potencia de cálculo, la sofisticación de los algoritmos y la disponibilidad de grandes volúmenes de información, la IA se ha integrado en muchos sectores, incluyendo la salud, la educación, la economía y el trabajo. No obstante, junto con sus beneficios en términos de eficiencia, innovación y mejora en la toma de decisiones, la IA también plantea importantes desafíos éticos, sociales y regulatorios, lo que hace necesario un análisis teórico que permita comprender sus fundamentos, alcances e implicaciones en la sociedad contemporánea.

La IA es un conjunto de capacidades cognitivas e intelectuales generadas por un computador y se crea con el propósito de construir dispositivos que puedan emular la capacidad de pensar y razonar de los seres humanos, y que, además, puedan perfeccionar el funcionamiento de acuerdo con la información que recopilen. La IA tiene un impacto en casi todas las áreas del conocimiento, como la medicina, la ciencia, la economía, la educación, la seguridad, entre otros (Delgado & Rodríguez, 2025).

Considerando el crecimiento acelerado de la IA a partir de factores tecnológicos clave como el aumento de la capacidad de procesamiento, el desarrollo de algoritmos avanzados y la disponibilidad de grandes volúmenes de datos, así como la idea de que la IA no se limita únicamente a la automatización de tareas, sino que implica la posibilidad de mejora continua, lo que amplía significativamente su campo de aplicación. La versatilidad y relevancia de IA en los pasillos de la medicina, la educación y la economía es digna de mención.

La IA es cada vez más relevante en la transformación de la sociedad, debido a su rápido desarrollo y a su capacidad para automatizar tareas y apoyar la toma de decisiones en diversos ámbitos. Esto permite comprender su alcance y utilidad en el contexto actual; sin embargo, el planteamiento podría fortalecerse si se incorporara una reflexión sobre los retos sociales y éticos que surgen con su implementación, a fin de ofrecer una visión más equilibrada del tema.

Díaz y Morales (2024) argumentan que este siglo, se ha constituido como una etapa de la historia en la que la tecnología ha tomado predominancia en cada uno de los aspectos de las personas, expandiéndose incluso al ámbito de las relaciones laborales. La aplicación de IA no es una práctica ajena para las empresas. Por el contrario, cada día su implementación está más presente. Incluye teorías, antecedentes, modelos conceptuales y definiciones clave.

Por último, Villagomez (2025) señala que la IA impacta innovadoramente en la forma de relacionarse con los entornos tecnológicos y en los diferentes modos de comunicarse, aprender y divertirse.

El coaching en las organizaciones: enfoques, liderazgo y efectividad

Desde la perspectiva organizacional y la acelerada transformación, la digitalización y la necesidad de adaptación constante, el *coaching* ha surgido como una herramienta estratégica para que el talento humano se desarrolle y fortalezca. Se ha consolidado como un proceso de acompañamiento orientado a potenciar las competencias individuales y colectivas, mejorando el desempeño laboral y promover climas organizacionales basados en la confianza, la reflexión y el aprendizaje continuo. En ese sentido, diversos enfoques tanto psicológicos como organizacionales, han expuesto su impacto positivo en la productividad, el bienestar, la inteligencia emocional y la efectividad directiva, así como su capacidad para hacer frente a los desafíos del trabajo híbrido y la integración de nuevas tecnologías, incluida la IA. En este sentido, el *coaching* se identifica como un elemento clave para la

transformación organizacional sostenible y la competitividad en entornos empresariales cada vez más complejos.

Según Del Toro et al. (2024), el término *coaching* proviene del francés y significa "guiar a una persona de un punto a otro", lo cual encaja de manera coherente con el significado actual de este proceso (Haas, 1992). Desde este punto de vista, se puede considerar un proceso, especialmente uno en el que el enfoque está en el desarrollo constructivo y activo del potencial que los seres humanos poseen dentro de las organizaciones.

Según Arboleda (2024), el *coaching* ha demostrado ser una herramienta esencial para el desarrollo personal y profesional, similar a los principios de la psicología positiva que reconoce las fortalezas del individuo, el bienestar general y el florecimiento humano positivo.

Los modelos psicológicos y organizacionales del *coaching* son útiles para análisis diferenciados, aunque en la práctica son interdependientes, dado que afectan directa y proporcionalmente a los resultados organizacionales. La necesidad que los líderes adquieran competencias propias del *coaching*, implica una evolución de los modelos de dirección tradicionales, en donde el directivo como aprendiz, se transforma en un facilitador del aprendizaje y crecimiento de manera individual y colectiva.

Las transformaciones de los estilos de liderazgo hacia el más transformacional se alinean de forma coherente con los cambios de los contextos organizacionales actuales, dado que promueve la creación de visiones comunes y sentimientos de pertenencia que favorecen la disposición y la corresponsabilidad de los integrantes de la organización. Por otra parte, el reconocimiento de los trabajadores como personas activas, con la disposición de aprendizaje, adaptación y de realizar aportaciones creativas, implica la necesidad de formular estrategias que favorezcan la construcción de climas laborales positivos que estimulen el desempeño productivo y que, además, fortalezcan la capacidad de las organizaciones para hacer frente a los nuevos escenarios.

El *coaching* es una herramienta que, además de ayudar en el crecimiento individual, puede impactar de manera eficiente el desarrollo de una organización en su conjunto, mejorando la sinergia del talento con los objetivos organizacionales a través de procesos reflexivos que mejoran las habilidades blandas y técnicas.

Podemos decir que en las organizaciones actuales, la complejidad y la competitividad son definitorias, y la creación de estrategias que permitan potenciar el liderazgo y trabajar

colaborativamente hacia la mejora del desempeño institucional son necesarias. El *coaching* gerencial, por lo tanto, se ha consolidado como un recurso que agrega valor estratégico y responde a la dinámica del mercado. Más allá de la creación de nuevas habilidades, el *coaching* gerencial promueve la resolución de problemas, la gestión del cambio y la cultura de la organización en el aprendizaje organizacional. La digitalización también ha transformado los enfoques tradicionales, lo cual implica el desarrollo de nuevas competencias tecnológicas, el uso de dispositivos virtuales y la movilidad en los escenarios laborales.

Vela y Flórez (2024) señalan que el *coaching* gerencial en las organizaciones de hoy busca desarrollar el talento a través de un liderazgo estratégico. Con *managers* y/o líderes *coach*, el proceso busca detectar y desarrollar todas las competencias y potencialidades de cada uno de los integrantes de los equipos, organizando esfuerzos alineados a los objetivos de la organización para aumentar la productividad y la rentabilidad.

El uso de la tecnología da los tipos de cambios que pueden acontecer en la gerencia. Manejar de forma correcta las nuevas tecnologías en la gerencia, da la posibilidad de extraer funciones a roles, junto a un desarrollo efectivo de los funcionamientos y las capacidades de los empleados, además de las habilidades de liderazgo.

La efectividad del *coaching* también se manifiesta de manera concreta en escenarios profesionales específicos. Lejos de considerarse una carga adicional, el *coaching* aparece como un elemento que alivia a largo plazo, la presión laboral, mostrando su potencial transformador en contextos complejos y altamente demandantes (Arboleda, 2024; Atencia & Yupanqui, 2024).

Como se describe, el debate sobre la efectividad del *coaching* en el ámbito organizacional ha ido evolucionando a la necesidad de evaluaciones más precisas y estructuradas que se relacionen con el impacto que el *coaching* tiene en el desarrollo organizacional.

Uso de la inteligencia artificial en el coaching empresarial

Como se ha descrito con anterioridad, la transformación digital mediante el uso de la IA ha comenzado a desempeñar un papel estratégico en múltiples disciplinas orientadas al desarrollo humano, entre ellas el *coaching*. Sin embargo, la necesidad de optimizar el desempeño individual y organizacional ha impulsado la adopción de soluciones tecnológicas capaces de analizar grandes volúmenes de información, identificar patrones de comportamiento y ofrecer retroalimentación personalizada.

El *coaching* asistido por IA emerge como una alternativa innovadora que complementa los enfoques tradicionales, permitiendo procesos más ágiles, precisos y orientados a resultados. Por lo que este apartado, analiza el impacto de la IA en el *coaching*, destacando sus beneficios en la efectividad de los programas de desarrollo, la mejora del desempeño y la generación de valor tanto para las personas como para las organizaciones.

Vorecol (2025) señala que la IA en el *coaching* se ha vuelto muy importante para profesionales y empresas que quieren mejorar el desarrollo personal y profesional. Con la tecnología en continua evolución, el *coaching* apoyado por IA está transformando la forma en que las personas y los grupos producen cambios en el comportamiento. En un informe de McKinsey se indica que el uso de IA en *coaching* podría disminuir hasta un 40% el tiempo requerido para lograr cambios significativos en hábitos y toma de decisiones.

El *coaching* sigue siendo fundamental; sin embargo, los sistemas de IA, al ser utilizados como una herramienta, potencializan el alcance, eficiencia e impacto del *coaching*, convirtiéndose en un activo primordial para las organizaciones que quieren lograr un impacto positivo en el contexto actual.

Materiales y método de investigación

Diseño de la investigación

Para llevar a cabo este análisis, apostamos por una revisión sistemática de literatura bajo el método Colin (2007), que nos pareció ideal para estructurar este tipo de investigaciones teóricas sin perder el rigor académico. Este enfoque permite rastrear, organizar y sintetizar de manera coherente lo que se ha escrito sobre IA y *coaching* empresarial, dos temas que, aunque parecen estar en mundos distintos, cada vez se cruzan más en la práctica organizacional.

Este método se enfoca en cinco pilares. Primero, la cobertura, que se refiere a qué tan amplia es la búsqueda (si se están tocando todos los rincones relacionados al tema). La segunda es la organización, que indica la estructuración y el orden de la información de tal forma que tenga sentido. La tercera es la lógica, que se refiere a la consistencia de los argumentos y los resultados. La cuarta es la integración, que trata de construir lazos entre distintos trabajos y enfoques. Y, por último, la narrativa, que es la relación que se puede contar la historia con claridad y sin tecnicismos, aunque sin caer en la superficialidad.

Estrategia de búsqueda y selección de fuentes

Se realizó la búsqueda de la literatura en el último trimestre de 2024 y en enero de 2025, priorizando publicaciones recientes (2020-2025) que permitan identificar las tendencias más recientes en el campo. Las bases de datos que se consultaron fueron *Scopus*, *Web of Science*, *Google Scholar* y repositorios institucionales en las áreas de administración, psicología organizacional y tecnología.

Hay términos de búsqueda que integran conceptos en inglés y en español tales como, *artificial intelligence*, *coaching*, *organizational development*, *leadership*, *digital transformation*, y en español, IA, desarrollo organizacional, y gestión del talento humano. Se usaron operadores booleanos (Y, O) para refinar los resultados y comprobar que las fuentes realmente abordaban la intersección de los dos temas.

Se consideraron artículos revisados por pares, capítulos de libros, reportes de investigaciones y disertaciones sobre el uso de la IA en las organizaciones, el *coaching* como mecanismo de desarrollo, o la interrelación de ambas. Se excluyeron trabajos de tipo técnico sobre algoritmos de IA sin ninguna vinculación organizacional y aquellos que se ocupan de *coaching* deportivo o *coaching* personal sin relación a la investigación.

Proceso de análisis y categorización

Una vez identificadas las fuentes relevantes, se procedió a un análisis de contenido temático en el que cada documento fue leído críticamente para extraer los argumentos centrales, evidencias empíricas (cuando las había), y marcos conceptuales utilizados. El análisis que realicé fue complejo. Involucró varias lecturas, comparaciones entre varios autores y la construcción de algunas categorías que capturaran tanto las coincidencias como las diferencias en el discurso académico.

Las categorías más relevantes derivadas del análisis apuntan a tres ejes en particular: 1) el alcance organizacional de la IA y sus fundamentos y los impactos organizacionales (beneficios y riesgos sociales/éticos); 2) *coaching* como estrategia organizacional y desarrollo empresarial y psicológico de la persona, la efectividad documentada, el liderazgo transformacional como un cruce y *coaching* como estrategia organizacional (psicológica); y, 3) la integración específica de la IA en *coaching* empresarial, en particular en modelos híbridos, competencias, plataformas digitales y nuevos elementos.

Se elaboraron tablas para cada categoría en donde se sintetizan los factores detectados, la cantidad de autores que los apoyaban, y las principales convergencias y divergencias teóricas. Esta actividad de síntesis, en un sentido más amplio, permitió apreciar no solo qué se ha dicho sobre estos aspectos, sino también cómo se elaboró dicho discurso, qué se dejó fuera, cuáles son los vacíos que se requieren ampliar y los temas que aún no son claros y requieren más atención.

Consideraciones éticas y limitaciones metodológicas

Como la investigación es de tipo teórica documental, las consideraciones éticas se basan principalmente en el cuidado de las fuentes: en el respeto a la autoría, en la citación adecuada, en la no tergiversación o en la mala contextualización de los argumentos de los autores. No se interactuó con participantes humanos ni se emplearon datos primarios.

Hay que tener en cuenta las limitaciones que tienen este tipo de estudios. En consecuencia, aunque se revisó de forma sistemática, podría darse el caso que fuentes relevantes queden fuera. Esto se puede deber a revisiones en otros idiomas, falta de acceso a ciertas bases de datos, o simplemente, a la inmensidad del campo. Por otra parte, la mayoría de los estudios encontrados son exploratorios o conceptuales, que poseen escasa evidencia empírica, lo que haya más difícil determinar conclusiones causales sobre el impacto de la IA en el *coaching* empresarial.

Otra limitación importante es el sesgo de publicación: la literatura académica tiende a reportar más casos de éxito que fracasos, lo que puede generar una visión excesivamente optimista sobre las tecnologías emergentes. Por ello, se intentó mantener una postura crítica y equilibrada al analizar los hallazgos, señalando tanto las promesas como las preocupaciones que rodean la integración de IA en procesos de desarrollo humano organizacional.

Análisis de resultados

La Tabla 1 sintetiza las variables analizadas: la primera columna indica la cantidad de elementos; la segunda presenta los factores identificados; la tercera muestra el número de autores que coincidieron en cada factor.

Tabla 1

Frecuencia de mención de los factores por autor

No.	Factores	Autor (es)	Cantidad de autor(es)
1	Escalabilidad y democratización del *coaching*	Terblanche; Passmore, Olafsson & Tee; Bernhardt & Rose; Barger; Diller; Stenzel & Passmore; Aboumoussa & Pfister; DataIntelo	7
2	Modelo híbrido como futuro óptimo	Terblanche; Passmore, Olafsson & Tee; Bernhardt & Rose; Bachkirova & Kemp; Aboumoussa & Pfister; International Coaching Federation	6
3	Limitaciones en inteligencia emocional	Barger; Passmore, Olafsson & Tee; Bachkirova & Kemp; Terblanche; Passmore & Tee; Aboumoussa & Pfister	6
4	Disponibilidad y personalización	Terblanche; Barger; Passmore, Olafsson & Tee; Bachkirova & Kemp; Aboumoussa & Pfister	5
5	Preocupaciones éticas	Diller, Stenzel & Passmore; Bachkirova & Kemp; Passmore, Olafsson & Tee; Aboumoussa & Pfister	4
6	Preferencia por coaches humanos	Barger; Passmore & Tee; Bachkirova & Kemp; Aboumoussa & Pfister	4
7	Reducción de costos	Terblanche; Passmore, Olafsson & Tee; Bernhardt & Rose	3
8	Resistencia de coaches profesionales	Diller, Stenzel & Passmore; Terblanche; Passmore, Olafsson & Tee	3
9	Efectividad en establecimiento de metas	Terblanche; Passmore, Olafsson & Tee; Barger	3
10	Liderazgo de *coaching* como moderador	Jeong, Kim & Lee; Jeong & Jeong	2
11	Uso de IA por coaches profesionales	International Coaching Federation; Vorecol	2
12	Aplicaciones de IA generativa en liderazgo	Aboumoussa & Pfister; Passmore, Olafsson & Tee	2

Tabla 1

Frecuencia de mención de los factores por autor

No.	Factores	Autor (es)	Cantidad de autor(es)
13	Evidencia empírica del efecto moderador del *coaching*	Jeong, Kim & Lee	1
14	Campo de investigación emergente	Passmore, Olafsson & Tee	1
15	Beneficios organizacionales del AI *coaching*	Aboumoussa & Pfister	1
16	Crecimiento de la industria del *coaching*	International Coaching Federation	1
17	Tamaño del mercado global	DataIntelo	1
18	Adopción empresarial de IA generativa	DataIntelo	1

La Tabla 2 presenta la síntesis de los factores identificados: la primera columna muestra el número de elementos, la segunda especifica los factores correspondientes y la tercera indica el número de autores que respaldaron cada factor.

Tabla 2

Factores críticos de éxito identificados en el estudio

No.	Factores	Cantidad de autor(es)
1	Terblanche; Passmore, Olafsson & Tee; Bernhardt & Rose; Barger; Diller; Stenzel & Passmore; Aboumoussa & Pfister; DataIntelo	7
2	Terblanche; Passmore, Olafsson & Tee; Bernhardt & Rose; Bachkirova & Kemp; Aboumoussa & Pfister; International Coaching Federation	6
3	Barger; Passmore, Olafsson & Tee; Bachkirova & Kemp; Terblanche; Passmore & Tee; Aboumoussa & Pfister	6
4	Terblanche; Barger; Passmore, Olafsson & Tee; Bachkirova & Kemp; Aboumoussa & Pfister	5
5	Diller, Stenzel & Passmore; Bachkirova & Kemp; Passmore, Olafsson & Tee; Aboumoussa & Pfister	4
6	Barger; Passmore & Tee; Bachkirova & Kemp; Aboumoussa & Pfister	4
7	Terblanche; Passmore, Olafsson & Tee; Bernhardt & Rose	3
8	Diller, Stenzel & Passmore; Terblanche; Passmore, Olafsson & Tee	3

Tabla 2

Factores críticos de éxito identificados en el estudio

No.	Factores	Cantidad de autor(es)
9	Terblanche; Passmore, Olafsson & Tee; Barger	3
10	Jeong, Kim & Lee; Jeong & Jeong	2
11	International Coaching Federation; Vorecol	2
12	Aboumoussa & Pfister; Passmore, Olafsson & Tee	2
13	Jeong, Kim & Lee	1
15	Passmore, Olafsson & Tee	1
16	Aboumoussa & Pfister	1
17	International Coaching Federation	1
18	DataIntelo	1

El uso de la IA en el *coaching* empresarial está determinado por múltiples factores. Predominan: Escalabilidad y democratización del *coaching*, Modelo híbrido como futuro óptimo, Limitaciones en inteligencia emocional, etc., totalizando diez variables principales.

Discusión y conclusiones

La presente revisión sistemática de literatura sobre el uso de la IA en el *coaching* empresarial reveló un campo de investigación emergente que se encuentra en un momento de transición crítica. La identificación y análisis de estudios en 18 factores permiten desgajar y comprender lo fundamental en esta intersección entre tecnologías y desarrollo humano organizacional.

Los autores mostraron que la democratización y escalabilidad del *coaching* son los hallazgos más relevantes, y se apoyan en siete autores. Este consenso indica que, en efecto, el *coaching* por IA tiene la mayor promesa en el *coaching* como una forma de incrementar la accesibilidad y o la disponibilidad de la práctica del *coaching* para ciertas poblaciones y organizaciones que en el pasado han estado en una situación de desventaja o marginación por cuestiones económicas o geográficas. La tecnología, en este sentido, no representa únicamente una herramienta de eficiencia, sino un vehículo de equidad organizacional.

El modelo híbrido es la segunda tendencia más mencionada por los investigadores, con seis citas. La integración de la IA con la intervención humana es, de forma más intuitiva,

la IA por su disponibilidad, procesamiento de datos, personalización, y el *coach* humano por su inteligencia emocional, y comprensión profunda, y espera el vínculo. Así, la tecnología no sustituye el *coaching* empresarial, sino que transforma roles y responsabilidades.

El hecho que seis autores señalen la falta de las emociones en la IA, es un hallazgo que resulta paradójico. Pese a haber avances en el procesamiento de lenguaje y reconocimiento de patrones de tipo emocional, existen aún limitaciones que se deben a la incapacidad técnica para el ámbito de la experiencia emocional.

Esto no es sólo una limitación técnica, sino que plantea una serie de preguntas filosóficas en relación con el acompañamiento profesional y los elementos que son difícil de resumir dentro del vínculo del *coach* y el *coachee* que van más allá de la transmisión de un contenido o la aplicación de un protocolo.

El poder, la flexibilidad y personalización que mencionan cinco autores son, sin duda, ventajas competitivas para las plataformas de *coaching* basado en IA. Poder realizar intervenciones de *coaching* personalizadas en cualquier lugar y en cualquier momento, sin duda, mejora la experiencia de los usuarios y, en consecuencia, la efectividad del desarrollo. Sin embargo, esta ventaja debe sopesarse con las preocupaciones de tipo ético que cuatro autores mencionan, sobre todo con la privacidad de los datos, sesgo algorítmico y falta de transparencia en la toma de decisiones automáticas.

Cuatro investigadores explican por qué los usuarios prefieren usar un *coach* físico en vez de un *coach* virtual. La gente que busca esta opción necesita poner su confianza en algo más seguro, aunque la tecnología avance, la gente siempre va a preferir un *coach* de *carne y hueso*. La gente necesita a alguien que les provea la *experiencia* que reclaman como sentimental y significativa, no un programa que solo cumpla con propósitos.

Los tres factores que construyen esta triada se complementan, cada uno desde la perspectiva de tres diferentes autores. La combinación de reducción de costos, la resistencia de los *coaches* profesionales, y la efectividad que se logre en el establecimiento de metas, dibuja las tensiones que, de manera casi permanente, atraviesan los procesos de transformación tecnológica en las profesiones más centradas en las relaciones humanas. La resistencia profesional, como parte de la reducción de costos, es una preocupación legítima ante la desvalorización de trabajos que requieren especialización y la posible pérdida de estándares. La efectividad en el establecimiento de metas, sugiere que la IA en la

personalización, la estructuración sistemática y el seguimiento riguroso, podrá concretar aportes significativos.

Los factores que los autores han mencionado en escasas ocasiones, no deben interpretarse como marginales. El *coaching* como liderazgo moderador, la IA utilizada por los coaches profesionales, la IA generativa en liderazgo y *coaching* y los *market indicators* son áreas de investigación potencial que en la medida que el campo se desarrolle, serán de relevancia.

Es importante reconocer que la revisión presenta limitaciones que son parte del estado del arte. La mayoría de los estudios que fueron analizados son de un tipo exploratorio o conceptual y tienen poca evidencia de un tipo experimental que permita de alguna manera establecer relaciones causales definitivas. Además, la parcialidad de las publicaciones en estudios de resultados, novedades positivas o en estudios que documenten fracasos o limitaciones, de manera significativa, sigue siendo relevante.

La adopción de la tecnología de *coaching* basada en IA plantea ciertas ineficiencias y resultados positivos de largo alcance. Asimismo, desde la ética, la cultura y la relación que define el éxito de estas intervenciones, la organización deberá considerar las ineficiencias. Desde un punto de vista tecnológico en *coaching*, la evidencia indica que el verdadero reto no es la tecnología, que será reemplazada en un instante, sino que el *coach* debe incorporar sistemas y competencias que empleen IA, sin deshumanizar el proceso.

Los avances por parte de los expertos simplemente demuestran que hay que construir sistemas que vayan más allá de la automatización de los procesos, que consideren la complejidad de la evolución humana y las relaciones que irremediablemente afectan la efectividad del *coaching*. El diseño de la experiencia del usuario, la transparencia en los algoritmos y las consideraciones éticas deben ser criterios fundamentales en la construcción de estas plataformas.

Las investigaciones futuras deberían enfocarse en el *coaching* basado en IA. Se requiere el establecimiento de estudios longitudinales, donde se documenten los efectos en el tiempo del *coaching* basado en IA. Además, se deberían hacer estudios de investigaciones comparativas, entre diferentes tipos de modalidades de intervenciones. Asimismo, el análisis de las variables que moderan los enfoques, de manera que se combinen los enfoques humanos, de tecnología, e híbridos. También es esencial el desarrollo de marcos éticos

específicos para el *coaching* asistido por IA, que rebase los principios generales de ética en IA y se ocupen de los efectos en el desarrollo organizacional humano.

Discusión

Los resultados obtenidos en esta revisión sistemática revelan un panorama complejo y multifacético sobre la integración de la IA en el *coaching* empresarial. La escalabilidad y democratización del *coaching*, identificada como el factor más prevalente con siete menciones, confirma la hipótesis planteada por Terblanche (2024) en cuanto a que la IA está redefiniendo, en esencia, el alcance del desarrollo organizacional. Este hallazgo dialoga directamente con las proyecciones de DataIntelo (2024) sobre el crecimiento acelerado del mercado de plataformas digitales de *coaching*, destacando que la accesibilidad de las soluciones ofrecidas, la cual, precisa el autor, resulta ser un motor de innovación tecnológica en este campo.

La figura del modelo híbrido con el rango de segunda categoría más citada (seis autores) genera, por tanto, una tensión productiva en el discurso hegemónico en construcción sobre la tecnologización del *coaching*. La obra de Passmore, Olafsson y Tee (2025), de Bernhardt y Rose (2024) y de Bachkirova y Kemp (2024) abogando por esta perspectiva integradora contradice las narrativas de sustitución tecnológica. La *International Coaching Federation* (2024) propone la coexistencia de lo humano y lo artificial. La Barger (2024) lo documenta, y sostiene que los sistemas de IA, aun cuando constituyan una alianza de trabajo, cualitativamente diferente a la de *coaches* humanos, son de este tipo. Esta diferenciación es importante en cuanto al modelo de alianza de trabajo de Bordin (1979) y los estudios de Graßmann, Schölmerich y Schermuly (2020) sobre *coaching*.

Las limitaciones en inteligencia emocional por seis autores deben ser interpretadas con cuidado. Recientes investigaciones como las de Elyoseph et al. (2023) sugieren que sistemas como *ChatGPT* pueden sobrepasar la evaluación de la conciencia emocional. Sin embargo, los resultados de esta revisión señalan la existencia de un escepticismo profesional. Esta contradicción se puede justificar por la diferencia entre el reconocimiento de patrones emocionales a nivel computacional, y la comprensión empática y contextual de la intervención humana. La teoría de la mente, de Damasio (1994), y las reflexiones de Baron-Cohen (1995) sobre el reconocimiento emocional, nos ayudan a entender por qué la evaluación computacional de las emociones no se compadece con la inteligencia emocional.

El tipo de preocupación ética que estos autores han descrito se puede relacionar con advertencias más generales sobre IA y, en particular, sobre Buolamwini y Gebru (2018), y los análisis de sesgos interseccionales en sistemas de clasificación automatizada. Diller, Stenzel y Passmore (2024) registran que, en su mayoría, las actitudes de los *coaches* profesionales hacia la IA se han incorporado de estas éticas, lo que indica que por el momento la tecnología se utilizará en estos campos no por su capital técnico, sino por la necesidad de crear regulaciones que protejan la privacidad, la justicia y la transparencia.

La resistencia profesional reportada por tres autores invita a reflexiones que trascienden el campo específico del *coaching*. Susskind y Susskind (2015) esperaron que las profesiones sofocadas por la transformación tecnológica seguirían estructuras de desagregación y recomposición, más que de simple automatización. Los resultados de esta revisión sugieren que el *coaching* podría estar en esta tanda, donde se desagregan y automatizan ciertas funciones (seguimiento de objetivos, provisión de recursos estructurados) mientras que otras se revalorizan como humanas.

Referencias

Aboumoussa, L., & Pfister, J. (2024). *Leadership development in the age of artificial intelligence* (Mossavar-Rahmani Center for Business & Government Associate Working Paper Series No. 244). Harvard Kennedy School. https://www.hks.harvard.edu/centers/mrcbg/publications/awp/awp244

Arboleda, V. A. (2024). Evidencia empírica y desafíos de la efectividad del coaching: una revisión sistemática. *Revista Académica Internacional de Educación Física*, *5*(3), 1-24. https://dialnet.unirioja.es/servlet/articulo?codigo=10173391

Atencia, D. I., & Yupanqui, M. D. (2024). *Influencia del coaching en el clima organizacional en una empresa privada de inversiones–Huánuco 2024* [Trabajo de grado, Universidad Continental]. Universidad Continental. https://app.compilatio.net/v5/report/5cf3e49918e7738f751a6adbcc95cc5c527af472/details

Bachkirova, T., & Kemp, R. (2024). 'AI coaching': Democratising coaching service or offering an ersatz? *Coaching: An International Journal of Theory, Research and Practice*, *17*(2), 129-143. https://doi.org/10.1080/17521882.2024.2368598

Barger, A. S. (2024). Artificial intelligence vs. human coaches: Examining the development of working alliance in a single session. *Front*. *Psychol.*, (15), 1364054. https://doi.org/10.3389/fpsyg.2024.1364054

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Bernhardt, M., & Rose, T. (2024, octubre 24). AI in coaching: A tool for the many, but not a replacement for the few. *Training Journal*. https://www.trainingjournal.com/articles/ai-coaching-tool-many-not-replacement-few

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy*: *Theory, Research & Practice*, *16*(3), 252-260. https://doi.org/10.1037/h0085885

Buolamwini, J., y Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, (81), 1-15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Colín, E.N. (2007). *Factores Organizacionales que Impactan en la Aplicación Exitosa de Células de Manufactura* [Trabajo de grado, Instituto Tecnológico de Cd. Juárez]. Repositorio del Instituto Tecnológico de Cd. Juárez, México.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

Del Toro Reyes, L., López Alfonzo , J. E., Rodríguez Durán, E. H., & Fernández García, J. F. (2024). Coach y Coaching como opción para contribuir al desarrollo individual y Organizacional. *GADE*: *Revista Científica*, *4*(4), 72-90. https://doi.org/10.63549/rg.v4i4.490

DataIntelo. (2024). *Digital coaching platform market research report 2024-2033: Global industry analysis, market size, share, trends, and forecast.* Dataintelo Consulting Pvt Ltd. https://dataintelo.com/report/digital-coaching-platform-market

Delgado Ramos, A., & Rodríguez Díaz, A. (2025). Clasificación, Aplicaciones y Desafíos de la Inteligencia Artificial: Un Análisis Integral. *Revista de Información científica para la Dirección en Salud*. https://revinfodir.sld.cu/index.php/infodir/article/view/1848

Díaz, L. D., & Morales, S. (2024). Impacto de la inteligencia artificial en las relaciones laborales. *Revista Digital Derecho del Trabajo y de la Seguridad Social*, 163-185.

Diller, S. J., Stenzel, L. C., & Passmore, J. (2024). The coach bots are coming: Exploring global coaches' attitudes and responses to the threat of AI coaching. *Human Resource Development International*, *27*(4), 597-621. https://doi.org/10.1080/13678688.2024.2375934

Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology*, (14), 1199058. https://doi.org/10.3389/fpsyg.2023.1199058

Graßmann, C., Schölmerich, F., & Schermuly, C. C. (2020). The relationship between working alliance and client outcomes in coaching: A meta-analysis. *Human Relations*, *73*(1), 35-58. https://doi.org/10.1177/0018726718819725

International Coaching Federation. (2024). Global coaching study: Coaching and technology – Innovating the future of coaching. *International Coaching Federation.* https://coachingfederation.org/research/global-coaching-study

Jeong, J., y Jeong, I. (2025). Fostering proactivity in the age of AI: Exploring the interplay of job reflection and coaching leadership. *Baltic Journal of Management*, *20*(1), 78-95. https://doi.org/10.1108/BJM-09-2024-0532

Jeong, J., Kim, B. J., & Lee, J. (2024). Navigating AI transitions: how coaching leadership buffers against job stress and protects employee physical health. *Frontiers in Public Health*, (12), 1343932. https://doi.org/10.3389/fpubh.2024.1343932

Passmore, J., & Tee, D. (2024). Human coaches and AI coaching agents: An exploratory quasi-experimental study of workplace client attitudes. *Journal of Work-Applied Management*. https://doi.org/10.1108/JWAM-02-2025-0032

Passmore, J., Olafsson, B., & Tee, D. (2025). A systematic literature review of artificial intelligence (AI) in coaching: Insights for future research and product development. *Journal of Work-Applied Management*, *17*(1), 1-28. https://doi.org/10.1108/JWAM-11-2024-0164

Vela, O. L., & Flórez, M. (2024). Coaching gerencial en la era moderna y su impacto en las organizaciones. *Revista de divulgación científica. Ciencia, Innovación, Tecnología, Ambiente y Sociedad*, *11*(1), 64-80. https://dialnet.unirioja.es/servlet/articulo?codigo=10329995

Villagomez Palacios, A. H. (2025). El impacto de la Inteligencia Artificial en la Sociedad: Una Revisión Sistemática de su Influencia en Ámbitos Sociales, Económicos y Tecnológicos. *Ciencia Latina Revista Científica Multidisciplinar*, *9*(1), 8150-8172. https://doi.org/10.37811/cl_rcm.v9i1.16468

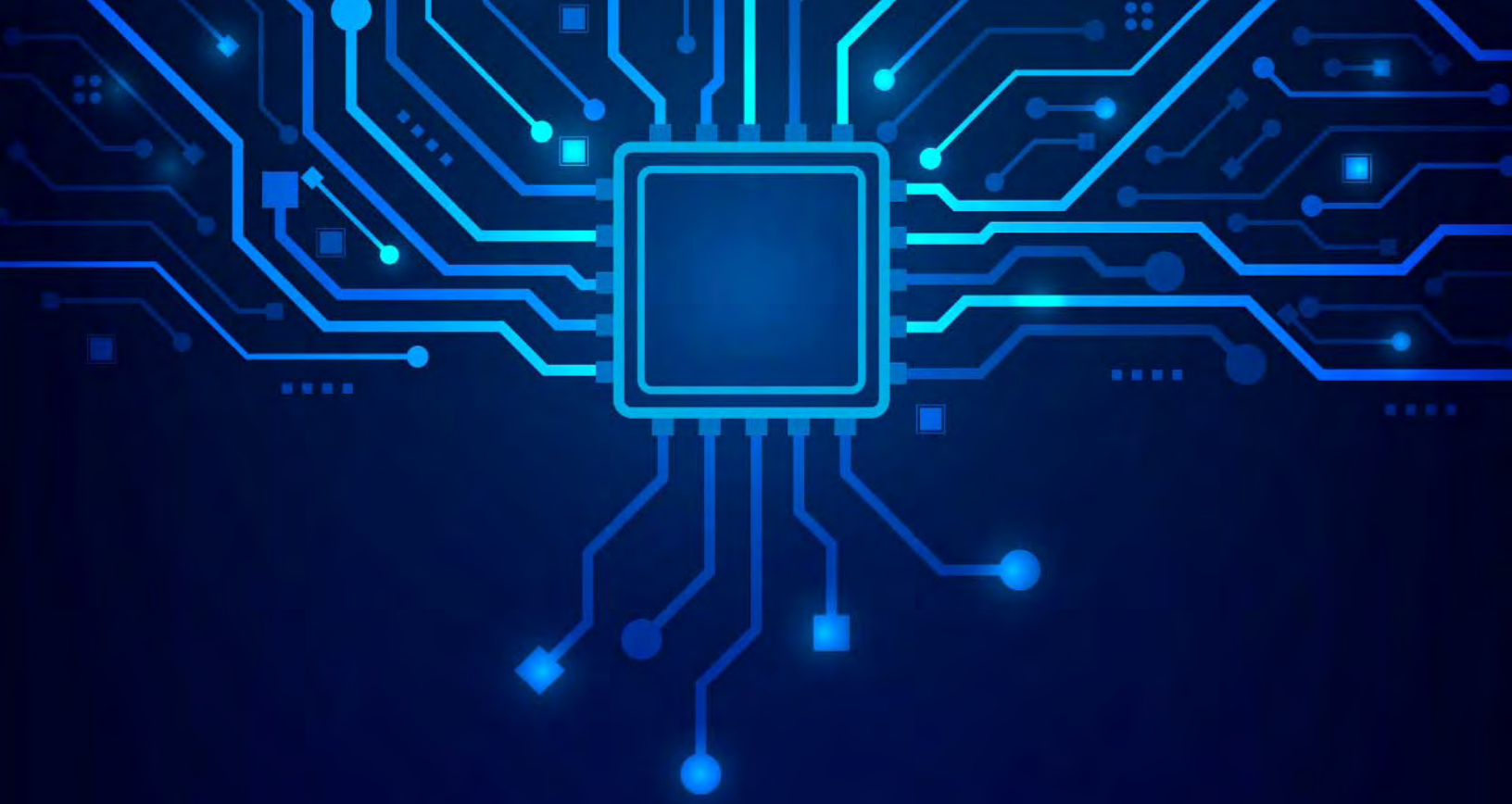
Vorecol, E. d. (2025). *Integración de inteligencia artificial en plataformas de coaching: beneficios y desafíos*. Vorecol. https://blogs-es.vorecol.com/articulo-integracion-de-inteligencia-artificial-en-plataformas-de-coaching-beneficios-y-desafios-169402?utm_source=chatgpt.com

Writing Group for the Women's Health Initiative Investigators. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *Jama*, *288*(3), 321-333. https://jamanetwork.com/journals/jama/fullarticle/195120/

Susskind, R.,& Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press.

Terblanche, N. H. D. (2024). Artificial Intelligence (AI) coaching: Redefining people development and organizational performance. *Journal of Applied Behavioral Science*, *60*(3), 412-435. https://doi.org/10.1177/00218863241283919

Vorecol. (14 de octubre de 2024). Integración de inteligencia artificial en plataformas de coaching y mentoring: Oportunidades y desafíos. *Blogs Es.vorecol*. https://vorecol.com/es/articulos/articulo-integracion-de-inteligencia-artificial-en-plataformas-de-coaching-y-mentoring-oportunidades-y-desafios-161762



10.

El uso de la inteligencia artificial en los procesos de investigación científica

Marysol Estrella Hernández García
Universidad Autónoma de Guerrero, México
ORCID: 0009-0001-6989-1142

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.10

10.

El uso de la inteligencia artificial en los procesos de investigación científica

Introducción

La investigación científica contemporánea se desarrolla en un entorno marcado por la digitalización y el uso intensivo de tecnologías computacionales, entre las cuales la inteligencia artificial (IA) ocupa un lugar cada vez más relevante. Su presencia en distintas fases del proceso de investigación —desde la recolección y análisis de datos hasta la formulación de hipótesis y la interpretación de resultados— no supone únicamente un avance técnico, sino una modificación en las condiciones bajo las cuales se produce y se valida el conocimiento científico.

Durante buena parte de la modernidad, la ciencia se entendió como una práctica guiada por el razonamiento humano, el juicio crítico y la deliberación metodológica, en la que el investigador asumía la responsabilidad epistémica de sus decisiones. Sin embargo, la incorporación progresiva de sistemas de IA ha comenzado a alterar este esquema, al introducir formas de mediación tecnológica que inciden directamente en la toma de decisiones científicas y tensionan los modelos tradicionales de responsabilidad (Floridi et al., 2018).

Estos sistemas ya no se limitan a automatizar tareas instrumentales. En diversos campos, participan en procesos sustantivos como la identificación de patrones, la jerarquización de variables y la generación de predicciones que orientan el rumbo de la investigación. Ello obliga a replantear quién decide, con base en qué criterios y bajo qué condiciones de control y responsabilidad.

Asumir que la IA opera de manera neutral dentro de la investigación resulta problemático. La evidencia muestra que los algoritmos incorporan supuestos, valores y sesgos derivados tanto de los datos de entrenamiento como de las decisiones humanas que intervienen en su diseño (O'Neil, 2016). La aceptación acrítica de sus resultados debilita el deber de justificación racional que ha sostenido históricamente la legitimidad del conocimiento científico.

El reconocimiento de estos riesgos no implica desconocer el potencial de la IA para fortalecer la investigación científica, sino delimitar las condiciones bajo las cuales su incorporación resulta epistemológica y jurídicamente legítima

Este capítulo examina el uso de la IA en la investigación científica desde una perspectiva crítica, ética y jurídica. Se trata de un ensayo documental orientado a analizar las transformaciones que la mediación algorítmica introduce en la producción del conocimiento y en la toma de decisiones científicas. El análisis se organiza en torno a cuatro ejes: la IA como herramienta metodológica, su incidencia en la decisión científica, sus implicaciones éticas y epistemológicas y la necesidad de marcos normativos y de gobernanza.

La integración de estos sistemas no puede entenderse como un asunto meramente técnico. La automatización de decisiones sin salvaguardas jurídicas compromete la responsabilidad científica y puede afectar la confianza social en la ciencia, lo que exige una reflexión crítica como la que se desarrolla en este trabajo.

La inteligencia artificial como herramienta en los procesos de investigación científica

La incorporación de sistemas de IA en los procesos de investigación científica no puede comprenderse únicamente como un avance técnico orientado a optimizar el análisis de datos. Desde un enfoque analítico, su utilización implica una transformación sustantiva en la forma en que se produce, valida y legitima el conocimiento científico. De ahí que, sostener que la IA actúa como una herramienta neutral dentro de la investigación resulta conceptualmente insuficiente y jurídicamente problemático, pues oculta los supuestos normativos, epistemológicos y valorativos que atraviesan su diseño y funcionamiento.

Tradicionalmente, la investigación científica se ha sustentado en la premisa de que el conocimiento es producido mediante procesos racionales controlables, atribuibles a sujetos responsables y sometidos a criterios de verificación, falsación y revisión crítica. Sin embargo, la introducción de sistemas algorítmicos capaces de procesar grandes volúmenes de información, detectar patrones y generar inferencias complejas reconfigura estas bases. La IA no se limita a acelerar tareas instrumentales, sino que interviene en fases sustantivas del proceso científico, influyendo en la delimitación del objeto de estudio, la selección de variables relevantes y la interpretación de resultados (Russell & Norvig, 2021).

Desde esta perspectiva, la caracterización de la IA como una simple herramienta de apoyo metodológico resulta reductiva. Aunque en su modalidad asistencial estos sistemas no sustituyen formalmente al investigador, sí introducen criterios de selección y jerarquización que condicionan el curso de la investigación. En consecuencia, el investigador ya no decide en un vacío epistemológico, sino dentro de un marco de recomendaciones algorítmicas que delimitan lo que aparece como relevante, eficiente o científicamente prometedor. Este desplazamiento, aunque sutil, tiene implicaciones profundas para la autonomía científica y para la responsabilidad asociada a la producción del conocimiento.

Uno de los principales riesgos de esta dinámica es la progresiva naturalización de la estandarización algorítmica. La promesa de eficiencia y reducción de errores humanos suele justificar la adopción de modelos que privilegian patrones dominantes y descartan anomalías, excepciones o enfoques no alineados con los parámetros del sistema. Bajo una mirada crítica, esta lógica puede conducir a una homogenización del pensamiento científico, limitando la pluralidad epistemológica que históricamente ha sido fuente de innovación y ruptura de paradigmas. En lugar de ampliar el horizonte del conocimiento, la IA corre el riesgo de reforzar enfoques ya consolidados, invisibilizando perspectivas críticas o marginales.

Este fenómeno resulta especialmente problemático si se considera que los sistemas de IA no operan de manera autónoma ni objetiva. Su funcionamiento depende de datos históricos, decisiones de diseño, criterios de optimización y supuestos implícitos definidos por desarrolladores e instituciones. Como advierte O'Neil (2016), los algoritmos pueden reproducir y amplificar sesgos existentes bajo una apariencia de neutralidad técnica. En el ámbito científico, donde la objetividad se asume como un valor central, esta reproducción de sesgos adquiere una gravedad particular, pues puede afectar la validez y legitimidad del conocimiento producido.

Desde el punto de vista jurídico, esta situación plantea interrogantes relevantes sobre la responsabilidad científica. Si los resultados de una investigación están mediados por sistemas algorítmicos que influyen de manera determinante en las decisiones metodológicas, Se vuelve indispensable cuestionar hasta qué punto el investigador conserva un control efectivo sobre el proceso. La delegación parcial de funciones cognitivas a sistemas automatizados puede diluir la atribución de responsabilidad, generando zonas grises difíciles de encuadrar dentro de los marcos tradicionales de imputación jurídica.

Asimismo, la dependencia creciente de herramientas de IA puede debilitar el ejercicio del juicio crítico por parte del investigador. La confianza excesiva en los resultados generados por sistemas automatizados tiende a desplazar la reflexión epistemológica y la revisión metodológica, sustituyéndolas por una aceptación acrítica de *outputs* algorítmicos. Este desplazamiento no solo afecta la calidad del conocimiento científico, sino que también compromete los principios de integridad y rigor que sustentan la práctica investigadora.

A partir de este análisis, el problema no radica en el uso de la IA en sí mismo, sino en la ausencia de límites claros que definan su función dentro del proceso científico. Cuando la IA deja de ser una herramienta controlada y se convierte en un actor decisional implícito, se produce una transformación estructural del proceso de investigación que no ha sido acompañada por una reflexión normativa equivalente. Esta asimetría entre innovación tecnológica y regulación jurídica genera un vacío que pone en riesgo la legitimidad del conocimiento producido.

Con ello, exige replantear el lugar de la inteligencia artificial dentro de la investigación científica desde una perspectiva jurídica y epistemológica integrada. Reconocer su potencial no implica aceptar acríticamente su expansión, sino delimitar los ámbitos en los que su uso resulta compatible con los principios de responsabilidad, control humano y transparencia que rigen la actividad científica. De lo contrario, la ciencia corre el riesgo de delegar funciones esenciales a sistemas cuya lógica de funcionamiento escapa al escrutinio crítico y a la rendición de cuentas.

Aceptar la IA como herramienta legítima en la investigación científica exige, por tanto, asumir que no se trata de un instrumento neutral ni inocuo. Su incorporación redefine las condiciones de producción del conocimiento y plantea desafíos que no pueden resolverse únicamente desde el plano técnico. Sin marcos normativos y criterios claros de control, la utilización de la IA puede erosionar los fundamentos epistemológicos y jurídicos que históricamente han otorgado legitimidad a la ciencia como práctica socialmente confiable.

La toma de decisiones mediada por inteligencia artificial en la investigación

La mediación de la IA en la toma de decisiones dentro de los procesos de investigación científica constituye uno de los desplazamientos más significativos —y menos problematizados— de la ciencia contemporánea. A diferencia de tecnologías previas, los sistemas

de IA no se limitan a ejecutar instrucciones predefinidas, sino que generan inferencias, recomendaciones y clasificaciones que orientan decisiones sustantivas en el desarrollo de la investigación. Esta capacidad transforma la relación entre el investigador y el proceso decisional, introduciendo un nuevo actor cuya intervención no siempre es visible ni jurídicamente delimitada.

En el ámbito científico, la toma de decisiones no es un acto meramente técnico, sino una práctica cargada de responsabilidad epistemológica y normativa. Decidir qué variables son relevantes, qué hipótesis merecen ser exploradas o qué resultados deben priorizarse implica juicios racionales que históricamente han sido atribuibles a sujetos identificables. Sin embargo, la incorporación de sistemas algorítmicos capaces de sugerir líneas de investigación *óptimas* o de jerarquizar datos en función de patrones estadísticos desplaza parcialmente estos juicios hacia una lógica automatizada (Floridi et al., 2018). Este desplazamiento no elimina la intervención humana, pero sí redefine sus márgenes de actuación.

En términos analíticos, el problema central no reside en la asistencia algorítmica, sino en la progresiva normalización de decisiones científicas mediadas por sistemas cuya lógica interna resulta opaca. Cuando un investigador acepta una recomendación algorítmica sin comprender plenamente los criterios que la sustentan, se produce una transferencia tácita de autoridad cognitiva. En estos casos, la decisión final sigue siendo formalmente humana, pero materialmente condicionada por un sistema cuya racionalidad no ha sido sometida a escrutinio científico ni jurídico.

Uno de los riesgos más relevantes de esta mediación es la incorporación de sesgos algorítmicos en el núcleo del proceso decisional. Los sistemas de IA aprenden a partir de datos históricos que pueden reflejar desigualdades estructurales, errores metodológicos o supuestos no cuestionados. Como señala O'Neil (2016), estos sesgos no solo se reproducen, sino que pueden amplificarse bajo la apariencia de objetividad técnica. En el contexto científico, esta situación resulta particularmente problemática, pues los resultados producidos mediante IA tienden a ser percibidos como más fiables que aquellos derivados exclusivamente del juicio humano.

La presencia de sesgos en la toma de decisiones científicas mediadas por IA compromete directamente la calidad del conocimiento producido. Cuando un sistema algorítmico prioriza determinadas correlaciones o descarta datos considerados *atípicos*, no solo influye en los resultados, sino también en la formulación misma de las conclusiones. En el plano

jurídico, esta dinámica plantea interrogantes sobre la validez de decisiones científicas que no pueden ser plenamente justificadas ni explicadas, lo que debilita los principios de transparencia y rendición de cuentas que deben regir la actividad investigadora.

Otro aspecto crítico es la opacidad de los sistemas de aprendizaje profundo utilizados en la investigación científica. Como advierte Burrell (2016), la dificultad para explicar cómo un algoritmo llega a determinadas conclusiones limita la posibilidad de someter sus resultados a procesos de verificación y replicación. Esta opacidad entra en tensión directa con uno de los pilares de la ciencia moderna: la posibilidad de cuestionar, reproducir y refutar los resultados obtenidos. Cuando las decisiones científicas se apoyan en sistemas que operan como *cajas negras*, se debilita la capacidad del propio sistema científico para autocontrolarse.

Desde el punto de vista jurídico, la opacidad algorítmica genera un problema adicional: la dilución de la responsabilidad. Si una decisión científica basada en IA conduce a conclusiones erróneas o a consecuencias negativas, resulta complejo determinar quién debe responder por ello. El investigador, el desarrollador del sistema, la institución que lo implementa o el proveedor de los datos aparecen como actores potencialmente responsables, sin que exista un marco normativo claro que permita imputar responsabilidades de manera efectiva. Esta dispersión erosiona los mecanismos tradicionales de control jurídico sobre la actividad científica.

En este contexto, el juicio humano adquiere una relevancia renovada, pero no en un sentido meramente simbólico. La deliberación ética, la interpretación contextual y la evaluación crítica de los resultados no pueden ser sustituidas por sistemas algorítmicos sin comprometer la legitimidad del conocimiento producido. La aceptación acrítica de recomendaciones generadas por IA implica una renuncia implícita al deber científico de justificación racional, transformando la toma de decisiones en un acto de validación automática más que en un ejercicio reflexivo.

Desde el plano teórico, la mediación algorítmica en la toma de decisiones científicas no debe entenderse como un proceso neutral ni inevitable. Su expansión responde a incentivos institucionales vinculados a la eficiencia, la productividad y la competitividad académica, que tienden a privilegiar resultados rápidos y cuantificables. Sin embargo, esta lógica puede entrar en conflicto con los valores que históricamente han sustentado la investigación científica, como el rigor, la prudencia y la responsabilidad frente a las consecuencias del conocimiento producido.

El desafío, por tanto, no consiste en excluir la IA de la toma de decisiones científicas, sino en delimitar normativamente su alcance. Hace imprescindible establecer criterios claros que definan qué decisiones pueden ser asistidas por sistemas algorítmicos y cuáles deben permanecer bajo control humano directo. Sin estos límites, la mediación algorítmica corre el riesgo de convertirse en una forma de automatización encubierta que desplace silenciosamente la responsabilidad científica.

Aceptar decisiones científicas mediadas por IA sin mecanismos adecuados de control y explicabilidad implica una transformación profunda del proceso investigador que no ha sido suficientemente debatida desde el derecho. Esta ausencia de problematización normativa resulta incompatible con la centralidad que la responsabilidad y la transparencia ocupan en la legitimación del conocimiento científico.

En última instancia, delegar decisiones sustantivas a sistemas opacos sin salvaguardas jurídicas adecuadas no representa un avance neutral, sino una reconfiguración del poder decisional en la ciencia que exige ser críticamente evaluada.

Implicaciones éticas y epistemológicas del uso de inteligencia artificial en la investigación científica

El uso de sistemas de IA en la investigación científica no solo introduce nuevas herramientas metodológicas, sino que plantea una transformación profunda en los fundamentos éticos y epistemológicos que sostienen la producción del conocimiento. A la luz de este planteamiento, la mediación algorítmica cuestiona de manera directa nociones clásicas como la objetividad, la neutralidad y la confiabilidad científica, al incorporar procesos decisionales cuya lógica no siempre es transparente ni atribuible a sujetos responsables.

Por ello, la ética y la epistemología dejan de ser marcos abstractos de reflexión para convertirse en ejes centrales de análisis jurídico.Uno de los supuestos más problemáticos en el discurso contemporáneo sobre IA es la idea de que los sistemas algorítmicos producen resultados objetivamente superiores a los del juicio humano.

Esta narrativa, ampliamente difundida en contextos científicos e institucionales, ignora que la IA no genera conocimiento en un vacío normativo. Por el contrario, sus resultados están condicionados por decisiones humanas previas relativas a la selección de datos, el diseño del modelo y los criterios de optimización. Como sostiene Floridi (2019), los sistemas

de IA actúan como agentes epistemológicos derivados, cuya capacidad de producción de conocimiento depende de estructuras normativas y valorativas preexistentes.

Desde una perspectiva epistemológica, esta mediación algorítmica tensiona la noción tradicional de objetividad científica. Si bien la ciencia nunca ha sido completamente neutral, la introducción de sistemas opacos refuerza una falsa apariencia de imparcialidad técnica. Los resultados generados por IA tienden a presentarse como verdades descontextualizadas, ocultando los supuestos que los sustentan.

Esta opacidad no solo dificulta la evaluación crítica del conocimiento producido, sino que debilita los mecanismos de corrección interna que han permitido históricamente el avance científico. La confiabilidad del conocimiento científico producido mediante inteligencia artificial depende, en gran medida, de la calidad, representatividad y contextualización de los datos utilizados para entrenar los modelos.

Sin embargo, en la práctica investigadora, estos elementos rara vez son objeto de escrutinio público o jurídico. En el marco de este análisis, aceptar resultados algorítmicos sin cuestionar las condiciones de su producción implica una renuncia implícita al deber científico de justificación racional. Esta renuncia resulta incompatible con los estándares epistemológicos que legitiman a la ciencia como práctica socialmente confiable.

En el plano ético, la mediación algorítmica introduce una problemática central: la redistribución de la responsabilidad científica. Cuando una decisión basada en IA conduce a interpretaciones erróneas, conclusiones sesgadas o consecuencias negativas, la atribución de responsabilidad se vuelve difusa. El investigador puede alegar que siguió recomendaciones del sistema; el desarrollador, que el modelo fue utilizado fuera de su contexto original; la institución, que adoptó una tecnología ampliamente aceptada.

Esta fragmentación de responsabilidades debilita los mecanismos tradicionales de rendición de cuentas y genera un vacío normativo difícil de subsanar. En términos normativos, esta dispersión resulta particularmente grave. La responsabilidad científica no es un mero principio ético, sino un elemento estructural que permite imputar consecuencias jurídicas cuando el conocimiento producido genera daños o afecta intereses protegidos.

Sin criterios claros que delimiten la responsabilidad en contextos de mediación algorítmica, la ética corre el riesgo de convertirse en un discurso retórico sin capacidad real de exigibilidad. Esto implica que, la ética de la IA no puede desvincularse de la construcción

de marcos normativos efectivos. La transparencia y la explicabilidad de los sistemas de IA aparecen, en este contexto, como requisitos indispensables para preservar la legitimidad del conocimiento científico.

No obstante, la insistencia en la explicabilidad técnica suele omitir una cuestión más profunda: la necesidad de justificar normativamente el uso de sistemas opacos en procesos decisionales relevantes. Como advierte Burrell (2016), la opacidad algorítmica no es solo un problema técnico, sino una forma de poder que limita la posibilidad de cuestionar y controlar los resultados producidos.

Considerando estos elementos, la aceptación de sistemas opacos en la investigación científica implica una transformación silenciosa de los estándares epistemológicos. Cuando la explicación de un resultado se sustituye por su mera eficacia predictiva, se produce un desplazamiento del conocimiento hacia una lógica instrumental que privilegia el resultado sobre la comprensión. Este desplazamiento no es neutro: redefine qué cuenta como conocimiento válido y quién tiene autoridad para producirlo.

Bajo esta lógica, la ética no debe concebirse como un complemento externo al desarrollo de la IA, sino como un elemento estructural de su integración en la investigación científica. Integrar principios éticos desde el diseño y la implementación de los sistemas de IA implica reconocer que ciertas formas de automatización resultan normativamente inadmisibles, aun cuando sean técnicamente eficientes. Sin este reconocimiento, la innovación tecnológica corre el riesgo de erosionar los valores que históricamente han sustentado la práctica científica.

Aceptar sin reservas la mediación algorítmica en la investigación científica implica asumir una concepción reducida del conocimiento, en la que la eficiencia y la predicción desplazan a la comprensión y la responsabilidad. Para el derecho, esta concepción resulta incompatible con la exigencia de transparencia, control humano y rendición de cuentas que deben regir toda actividad con impacto social. La ética de la IA, sin mecanismos normativos que la respalden, se convierte en una narrativa legitimadora que encubre decisiones estructuralmente problemáticas.

En consecuencia, las implicaciones éticas y epistemológicas del uso de la IA en la investigación científica no pueden resolverse mediante declaraciones de principios o códigos voluntarios. Requieren una reflexión jurídica profunda que permita delimitar responsabilidades,

establecer límites a la automatización y garantizar que la producción de conocimiento no se desvincule de los valores que le otorgan legitimidad. Sin estas salvaguardas, la ciencia corre el riesgo de avanzar tecnológicamente mientras retrocede en términos de responsabilidad y control normativo.

La necesidad de marcos normativos y de gobernanza para el uso de inteligencia artificial en la investigación

A pesar de la creciente incorporación de sistemas de IA en los procesos de investigación científica, los marcos normativos que regulan su uso siguen siendo fragmentarios, indirectos y, en muchos casos, insuficientes. Esta ausencia regulatoria no es neutra ni accidental, sino el resultado de una priorización normativa que ha privilegiado los usos comerciales y de mercado de la IA, relegando a un segundo plano su impacto en la producción del conocimiento científico. Esta omisión resulta especialmente problemática, pues la investigación científica constituye un ámbito con efectos estructurales sobre la toma de decisiones públicas, la formulación de políticas y la legitimación social del conocimiento.

En términos generales, los instrumentos jurídicos existentes no abordan de manera específica la mediación algorítmica en la investigación científica. Las normas sobre protección de datos, propiedad intelectual o ética en la investigación resultan insuficientes para regular sistemas que influyen activamente en decisiones metodológicas, interpretativas y conclusivas. Esta carencia genera un vacío normativo que dificulta la atribución de responsabilidades y debilita la protección de principios fundamentales como la integridad científica, la transparencia y el control humano efectivo.

Bajo este enfoque, el problema no radica únicamente en la falta de normas específicas, sino en la ausencia de una concepción jurídica clara sobre el papel de la IA en la ciencia. La regulación existente tiende a tratar a la IA como una herramienta técnica neutral, sin reconocer su capacidad de intervenir en procesos decisionales sustantivos. Esta concepción resulta jurídicamente insuficiente, pues impide identificar los riesgos estructurales asociados a la delegación de funciones cognitivas en sistemas automatizados y limita la posibilidad de establecer obligaciones claras para los distintos actores involucrados.

La omisión regulatoria adquiere mayor relevancia si se considera el papel de las instituciones públicas y de los organismos financiadores de la investigación. En muchos casos, la adopción de sistemas de IA responde a políticas institucionales orientadas a la eficiencia,

la productividad y la competitividad académica, sin que existan mecanismos normativos que exijan evaluaciones previas de impacto ético o jurídico. Desde esta perspectiva, la falta de regulación no solo es atribuible al legislador, sino también a las autoridades científicas y administrativas que han normalizado el uso de IA sin establecer salvaguardas adecuadas.

En el plano internacional, diversos organismos han propuesto principios orientadores para el desarrollo y uso responsable de la IA. No obstante, estos instrumentos suelen adoptar la forma de recomendaciones o declaraciones de carácter no vinculante, cuya eficacia depende de la voluntad de los Estados y de las instituciones científicas. Si bien estos marcos contribuyen a visibilizar los riesgos asociados a la IA, resultan insuficientes para garantizar la exigibilidad de responsabilidades en contextos concretos de investigación científica.

En el ámbito regulatorio, la ausencia de mecanismos vinculantes en materia de gobernanza de la IA limita la capacidad del derecho para intervenir de manera efectiva en la regulación de la ciencia algorítmica. La ética, sin respaldo normativo, se convierte en un discurso aspiracional que carece de fuerza coercitiva. Esta situación resulta incompatible con la relevancia social de la investigación científica y con el impacto que sus resultados pueden tener en la toma de decisiones públicas y privadas.

En el contexto europeo, los esfuerzos regulatorios basados en un enfoque de riesgo ofrecen elementos relevantes para repensar la gobernanza de la IA en la investigación científica. Sin embargo, incluso estos marcos presentan limitaciones cuando se trata de regular procesos epistémicos complejos. La clasificación de riesgos suele centrarse en los efectos externos de los sistemas de IA, dejando en segundo plano su impacto interno en la producción del conocimiento. Desde una perspectiva crítica, esta omisión reproduce la idea de que la ciencia es un ámbito autónomo ajeno a la regulación, lo cual resulta cada vez menos sostenible.

En el caso de contextos como el mexicano y latinoamericano, la ausencia de marcos normativos específicos se ve agravada por debilidades institucionales y asimetrías en el acceso a tecnologías avanzadas. La adopción acrítica de sistemas desarrollados en otros contextos normativos puede profundizar dependencias tecnológicas y reproducir modelos de investigación que no responden a las necesidades locales. Desde esta perspectiva, la falta de regulación no solo afecta la integridad científica, sino también la soberanía epistémica y la capacidad de los Estados para orientar la producción de conocimiento en función de intereses públicos.

Frente a este escenario, obliga a avanzar hacia modelos de gobernanza de la IA que reconozcan explícitamente las particularidades de la investigación científica. Más que imponer normas rígidas, se trata de establecer principios jurídicos exigibles que delimiten el uso legítimo de la IA en la ciencia. Entre estos principios destacan la responsabilidad clara y atribuible, la transparencia metodológica, el control humano significativo y la posibilidad de auditoría de los sistemas utilizados.

No obstante, cualquier propuesta regulatoria debe evitar una hiperlegalización que inhiba la creatividad científica. El desafío consiste en construir un marco normativo que funcione como límite y no como obstáculo, estableciendo criterios mínimos de legitimidad sin sofocar la innovación. Desde una lectura crítica, la ausencia total de regulación resulta tan problemática como una regulación excesiva, pues ambas pueden afectar negativamente la calidad y la legitimidad del conocimiento producido.

Aceptar la incorporación de la IA en la investigación científica sin marcos normativos claros implica una renuncia implícita a la función garantista del derecho. La ciencia no puede quedar al margen del control jurídico cuando sus procesos y resultados tienen efectos sociales relevantes. A partir de lo anterior, la construcción de modelos de gobernanza adecuados no es una opción discrecional, sino una exigencia derivada del compromiso del derecho con la protección de principios fundamentales y con la rendición de cuentas en la producción del conocimiento.

Reflexiones finales sobre el uso responsable de la inteligencia artificial en la investigación científica

La incorporación de la IA en los procesos de investigación científica constituye una de las transformaciones más profundas de la ciencia contemporánea, no por su capacidad técnica, sino por su impacto estructural en la producción y validación del conocimiento. A lo largo de este capítulo se ha sostenido que la IA no puede entenderse como una herramienta neutral ni meramente instrumental, pues su uso introduce criterios de selección, jerarquización y decisión que inciden directamente en el rumbo de la investigación científica y en la legitimidad de sus resultados.

Desde esta perspectiva, aceptar la integración de sistemas de IA en la investigación sin una problematización crítica implica asumir, de manera implícita, una concepción reducida del conocimiento científico, centrada en la eficiencia y la optimización, pero desvinculada de

la responsabilidad epistemológica y jurídica. La promesa de rapidez y capacidad de procesamiento no puede justificar una delegación acrítica de decisiones que históricamente han sido atribuibles a sujetos responsables y sometidos a estándares de justificación racional.

Retomar la tesis central desarrollada en este trabajo obliga a afirmar que el uso de IA en la investigación científica plantea desafíos éticos, epistemológicos y normativos que no pueden resolverse mediante declaraciones de principios ni mediante la confianza ciega en la autorregulación tecnológica. La mediación algorítmica reconfigura los criterios de objetividad, debilita la trazabilidad de las decisiones científicas y dificulta la imputación de responsabilidades cuando los resultados producen efectos socialmente relevantes. Ignorar estas tensiones no solo empobrece el debate académico, sino que compromete la función social de la ciencia como fuente de conocimiento legítimo.

El reto, por tanto, no consiste en rechazar la inteligencia artificial ni en frenar la innovación tecnológica, sino en establecer límites claros a su incorporación en la investigación científica. Integrar la IA de manera responsable exige reconocer que no todas las funciones cognitivas son normativamente delegables y que ciertas decisiones deben permanecer bajo control humano efectivo. La ausencia de estos límites transforma la asistencia algorítmica en una forma de automatización encubierta incompatible con los principios de responsabilidad y rendición de cuentas que sustentan la práctica científica.

En este sentido, la ciencia del futuro no dependerá únicamente del desarrollo de algoritmos más sofisticados, sino de la capacidad de las instituciones científicas y de los marcos jurídicos para regular su uso de manera crítica. Sin mecanismos normativos que garanticen la transparencia, la explicabilidad y la atribución clara de responsabilidades, la inteligencia artificial corre el riesgo de erosionar la confianza social en la ciencia y de consolidar prácticas investigadoras opacas difícilmente justificables desde el derecho.

Como líneas futuras de reflexión, resulta indispensable avanzar hacia enfoques interdisciplinarios que no se limiten a describir los beneficios técnicos de la inteligencia artificial, sino que analicen sus implicaciones estructurales en la producción del conocimiento. Asimismo, se vuelve urgente el diseño de instrumentos normativos específicos para la investigación científica que superen la lógica del *soft law* y establezcan criterios exigibles para el uso legítimo de sistemas algorítmicos. Sin estos instrumentos, la ética de la IA permanecerá en el plano declarativo, sin capacidad real de incidir en las prácticas científicas.

Aceptar la automatización de procesos decisionales sin estas salvaguardas no representa un avance neutral, sino una regresión en términos de responsabilidad científica y control jurídico. Desde una perspectiva crítica, la integración de la IA en la investigación solo puede considerarse legítima si se encuentra subordinada a principios normativos claros que preserven el juicio humano, la transparencia y la rendición de cuentas. De lo contrario, la innovación tecnológica corre el riesgo de avanzar a costa de los valores fundamentales que históricamente han otorgado legitimidad a la actividad científica.

Referencias

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

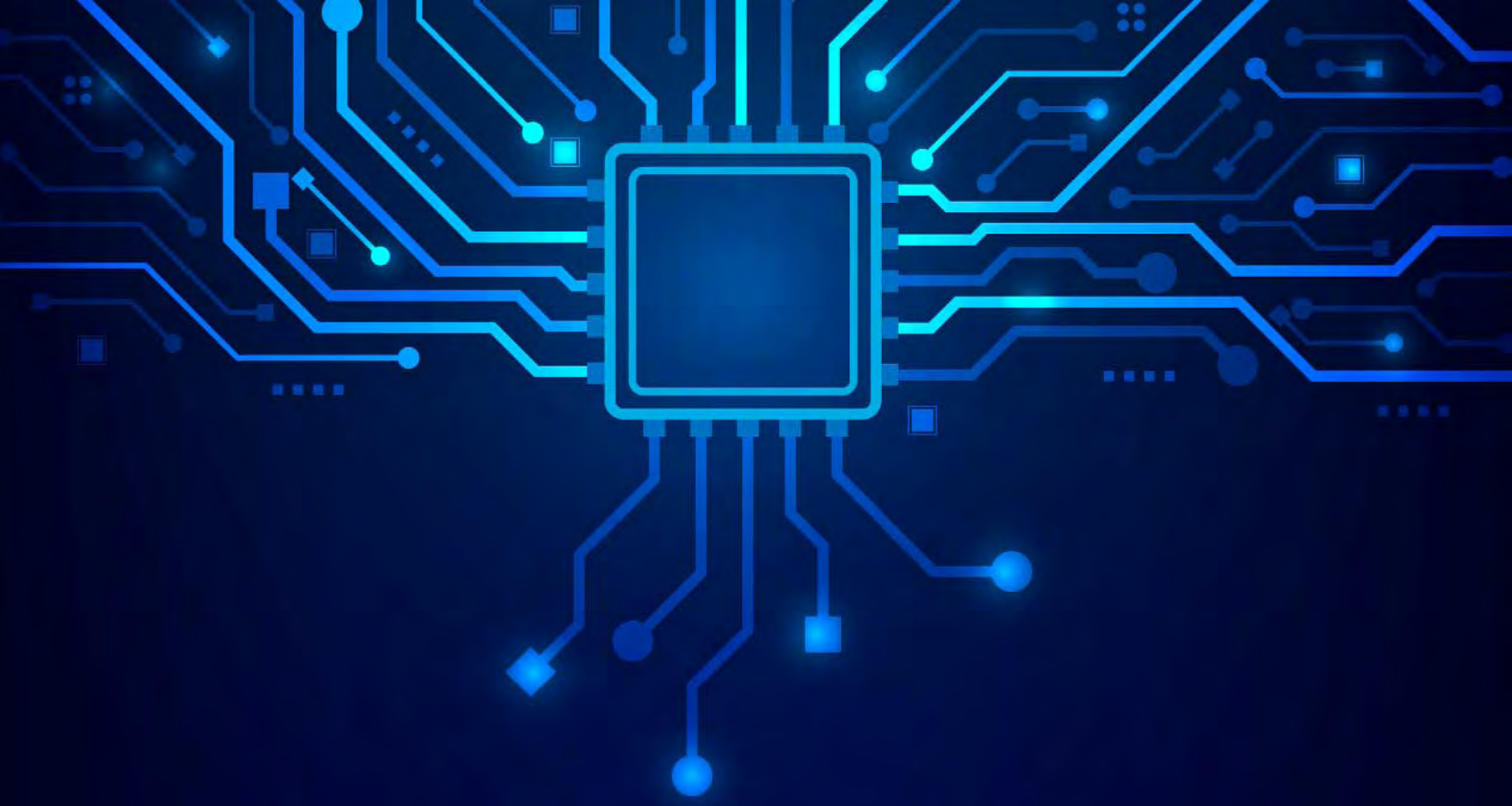
OECD. (2019). *OECD principles on artificial intelligence*. OECD. https://www.oecd.org/going-digital/ai/principles/

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown Publishing Group.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems*, *10*(4), 1–31. https://doi.org/10.1145/3419764

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137



11.

Hacia una nueva praxis de ciencia abierta dominada por datos masivos e inteligencia artificial generativa

Erik Carbajal-Degante
Universidad Nacional Autónoma de México
ORCID: 0000-0002-1532-9322

Leonardo Ledesma-Domínguez
Tecnológico de Monterrey, México
ORCID: 0000-0002-5374-3954

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.11

11.

Hacia una nueva praxis de ciencia abierta dominada por datos masivos e inteligencia artificial generativa

Introducción

Desde sus inicios, la investigación científica se ha posicionado como un elemento sustancial para el descubrimiento y consolidación de nuevo conocimiento. Sin embargo, la oportunidad de reinventarse siempre ha estado latente al sospechar que podría haber un cambio, ya sea paulatino o súbito de la forma en como realizamos investigación. Todos los procesos internos del llamado método científico han sucumbido a modificaciones producto de la necesidad social de cada época, el panorama de avance tecnológico, y los marcos epistemológicos dominantes. La investigación ha dado lugar a incorporar nuevas herramientas y lenguajes que reconfiguran los roles de los involucrados quienes persiguen un conjunto de objetivos e hipótesis para ayudar a interpretar la realidad actual. De entre las diversas disciplinas, las ciencias de la computación proveen de diversos enfoques que suman un impacto transformador al proceso de análisis-síntesis que envuelve el clásico marco metodológico de una investigación, lo que brinda una visión interdisciplinar en la adopción de nuevas herramientas para el estudio en cualquier otra disciplina de las ciencias.

Ante este escenario inminente de cambio, la disrupción propiciada por la manera inesperada en la que emergen nuevas herramientas con fines de estudiar las problemáticas del contexto actual, irrumpen y alteran las prácticas consolidadas. Este ritmo dinámico no siempre ha sido anticipado por la comunidad científica, por lo que generan incomodidades y debates sobre su verdadero impacto al proponer cambios paradigmáticos incluso sobre las bases de conocimiento ya existentes. Es precisamente la tensión derivada del uso de herramientas innovadoras que genera discrepancias y cuestionamientos a años de ciencia con resultados y metodologías que parecían incuestionables.

Dentro de los elementos se destacan las aportaciones traídas por las ciencias de la computación, como la propia computadora e internet, cuyo objetivo en investigación siem-

pre ha sido el robustecer el proceso metodológico en todas sus fases, lo que ha llevado a facilitar ciertas tareas en sentido automatizable, optimizable y computable. A esto se le suman los elementos de recién emersión, como los datos masivos y la inteligencia artificial generativa (IAGen), que proveen una estrategia de análisis de datos a grandes escalas, y en mayor medida, las capacidades de generación de nuevos datos, difuminando las fronteras entre lo automatizable y lo autónomo, lo real y lo artificial, lo humano y lo no humano.

Ante este panorama, el presente capítulo tiene el propósito de analizar el papel que desempeñan estos nuevos elementos en la investigación científica contemporánea. Lejos de adoptar una postura tecno-determinista, se propone una reflexión que examine argumentativamente el impacto asociado a las nuevas prácticas metodológicas y contribuir a la comprensión de la disruptividad en la reconfiguración de la producción de nuevo conocimiento, con miras en un cambio encabezado por estos elementos hacia una nueva etapa de ciencia abierta.

La relevancia de observar

Hasta responder la pregunta más sencilla podría traer consecuencias irreversibles a nuestras vidas, tal como el vuelo de una mariposa podría desencadenar situaciones que impactan en las condiciones de un sistema físico, mecánico e incluso social. Mediante esta analogía, Lorenz (1963) en su teoría del caos establece que sistemas, al parecer deterministas (resultado inevitable de eventos anteriores), podrían ser caóticos a nuestra forma de ver, debido a que no conocemos todas las variables que influyen en la manera de cómo observamos y medimos un evento.

Concretamente, si pudiéramos tener acceso a conocer todos los elementos que propician una causa, podríamos predecir con facilidad y certeza el futuro de una acción e incluso las determinaciones que construyeron su pasado. El futuro y el pasado se posicionarían ante nuestros ojos y serían tan claros como una ecuación matemática resoluble a pasos. Precisamente, LaPlace, en 1814, formuló un ensayo filosófico sobre la probabilidad, y como resultado, el experimento mental conocido como el demonio de LaPlace (1814), que tiene repercusiones en la ciencia, incluso hoy en día. Se trata de la utopía del determinismo científico que establece la posibilidad de predecir todo si, con detenimiento, observamos, analizamos y calculamos sobre toda la información, en caso de tener acceso a ella. De este modo, el demonio de LaPlace enfrenta un problema práctico: la imposibilidad de recopilar toda la información, y en este sentido, la causalidad deja de ser una cadena rígida para con-

vertirse en una red de probabilidades. Así, el indeterminismo abre una grieta en el enfoque del determinismo clásico y reconfigura la discusión sobre el libre albedrío, desplazando el debate desde la omnisciencia imposible hacia la estructura misma del ser.

De aquí se desprenden algunas interrogantes sobre el libre albedrío tales como: ¿realmente nuestras decisiones propician nuevos eventos? o bien, ¿y si nuestras decisiones son producto de decisiones previas? Precisamente este razonamiento contrapone al determinismo de tal forma que, si uno es verdadero, el otro podría ser falso; es decir, si el futuro no está completamente escrito, pudiese haber una salida a toda la cadena de causalidad. Si bien es cierto que estas dos perspectivas parecen contrastantes y suelen ser profundamente analizadas debido a su incompatibilidad, el determinismo blando, del filósofo David Hume (1748); sugiere redefinir el concepto de libre albedrío, y con ello argumentar que un acto es libre cuando sus causas provienen de uno mismo, más no cuando no tiene causa en sí. Es como llevar a cabo una de las primeras etapas de introspección con el que pretendemos dar sentido a un evento o situación. Esto representa un abordaje consolidado del pensamiento humano que se encuentra de manera muy frecuente en la filosofía e incluso en las ciencias.

La mirada de introspección desempeña un papel fundamental en el análisis del individuo; es producto directo de la conciencia (o lo que entendemos de ella como el atributo de estar y sentirse vivos) e impulsado por la observación, que nos ayuda a construir un marco de significancia al relacionar nuestra existencia con el universo, así como contemplarlo. De manera similar, a través de su contraparte, la extrospección, dirigimos nuestra atención al exterior y encontramos sentido de nuestras causas con un fenómeno externo, aunque no siempre hallamos una relación explícita. Ambos enfoques poseen rasgos en común que le permiten reducir sus atributos a una fuente importante de percepción y quizás la más utilizada por el ser humano, la cual se define como la observabilidad.

En un sentido amplio, mediante la observabilidad tenemos la capacidad de percibir un fenómeno a través de sus manifestaciones. Se basa en el principio de que podemos estudiar cualquier evento si miramos en la dirección correcta, con los instrumentos correctos. La observabilidad ha sido partícipe en posicionarnos como civilización y es en gran medida por la manera en cómo la hemos acoplado al momento de hacer ciencia, en cualquiera de sus enfoques y especialmente a través del método científico.

La clásica disputa entre empirismo y racionalismo

No fue sino hasta el siglo XVII que se estableció un marco conceptual robusto sobre la manera de realizar ciencia. Un enfoque científico particular es traído por el empirismo, que justamente resalta las capacidades de observar y experimentar del ser humano. Si observamos, podemos desarrollar un conjunto de habilidades complementarias (o sentidos) que nos ayuden a experimentar, y con ello a aprender. Sostiene que el conocimiento se construye a partir de la observación para generalizar y formalizar un compendio de hechos. Por ejemplo, en la disciplina llamada inteligencia artificial (IA), se sabe que el éxito de diversas herramientas del aprendizaje profundo (definido como un subcampo del aprendizaje de máquina) se debe, en gran medida, a los datos que recibe y las formas de entrenarse. Un enfoque del empirismo hacia este campo sería argumentar que los sistemas de IA funcionan correctamente gracias al entrenamiento de sus datos, algo muy similar a afirmar que la experiencia sensorial es fuente de conocimiento, lo cual es parcialmente cierto.

De hecho, existe un paradigma del aprendizaje de máquina conocido como aprendizaje por refuerzo, donde un robot (o sistema de IA) aprende a través de mecanismos de prueba-error, recompensas y objetivos (Sutton & Barto, 2020). Sin embargo, no todos los sistemas de IA tienen la capacidad de aprender de manera heurística, quizás otros solamente requieren un buen diseño independientemente del enfoque de entrenamiento o los datos que se les asigne. Digamos que es un sistema que por naturaleza es bueno al realizar una tarea.

La idea de ejecutar una acción perfecta de manera natural también podría ser vinculada a la deducción lógica, al análisis y a la razón, es decir, podemos alcanzar ciertas verdades que por lo general sucumben a nuestros sentidos, pero no a nuestra capacidad de razonar, que se posiciona como un principio fundamental dotado por la existencia, cuyo máximo exponente ha sido (Descartes, 1637). De hecho, la corriente del racionalismo sostiene (a diferencia del empirismo) que la razón es la fuente principal del verdadero conocimiento y la existencia de ideas innatas no provenientes completamente de la experiencia. Nos dan la pauta de cómo maniobrar al hacer ciencia, algo así como un elemento cercano a la intuición. Similar al caso empírico de sistemas de IA, una visión racionalista podría argumentar que ciertos sistemas de IA son naturalmente buenos realizando una tarea debido a su diseño matemático, que les permite desempeñarse correctamente sin depender tanto de otras variables o del proceso de entrenamiento en sí. Esta idea es comparable a la afirmación de que los sentidos pueden ser engañados, mientras que la facultad racional no lo es, una tesis que, si bien resulta sugerente, es parcialmente válida.

Desarrollo

El método científico y la inteligencia artificial generativa

Como tal, la dicotomía entre el empirismo y el racionalismo promueve un anclaje particular al momento de hacer ciencia si se adopta un enfoque drástico, por lo que el resultado podría resultar ambiguo. Por ejemplo, en el caso de abordar un estudio desde la perspectiva empírica sin mirada racional, se promueve una carencia interpretativa donde los hallazgos podrían parecer no tener sentido, lo que puede resultar en una especie de cientificismo. Por otro lado, el racionalismo sin empirismo promueve un abordaje teórico imposible de refutar, fomentando la especulación donde las ideas se desconectan de la realidad y pierden su relación con lo analíticamente observable. Esto da lugar a una especie de dogmatismo.

Con relación a la IA, hemos escuchado diversos argumentos que podrían caer en alguna de estas posturas tan contrastantes y en cierto sentido extremas; incluso, provenientes desde las mismas instituciones de educación superior. Por ejemplo, la prohibición en el uso de los modelos masivos de lenguaje (LLM) durante el 2023, que algunas instituciones educativas llegaron a impulsar (Johnson, 2023). Este enfoque abordaba la premisa de que “Hasta que un estudio científico no demuestre con absoluta certeza que una IA como ChatGPT no puede generar contenido peligroso, debemos prohibirla por completo”. En esta idea, se exige una prueba fehaciente de un negativo que es difícil de alcanzar, incluso hoy en día, mientras que se paraliza cualquier acción progresiva basada en un sentido de precaución tampoco comprobable.

Por otro lado, también hemos escuchado sobre la idea de que “La IA nunca podrá ser creativa o tener conciencia porque es simplemente matemáticas y código; carece de un alma o espíritu”. Una afirmación que parte de una definición esencialista de la naturaleza humana, posicionando al ser humano como elemento que inspira la creación de nuevas herramientas a su imagen. Esto representa una visión antropocentrista que cierra la puerta a nuevas investigaciones debido a la afirmación incuestionable de la naturaleza humana como única forma de conciencia.

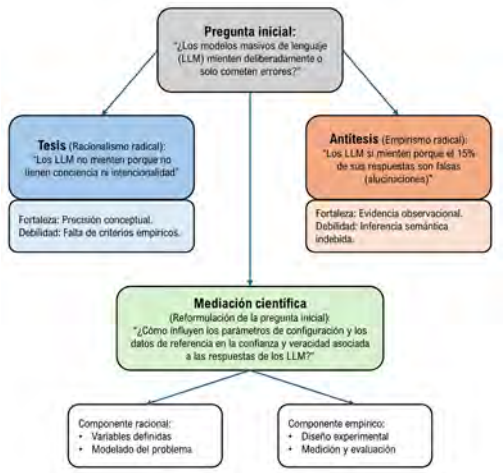
Desde luego, el método científico como lo conocemos trata de alinear las potencialidades de cada corriente, ayudando a complementar sus beneficios al converger en una metodología establecida que ha sido adoptada por años. Se sabe que, por un lado, el abordaje empírico provee de elementos sustanciales como la objetividad, la verificabilidad, la replicabilidad, entre otros, así como resalta las ventajas de añadir un proceso fundamental

de experimentación. Mientras que una estructura lógica, metodológica de generalización y modelado es provista por el racionalismo. De esta forma, el método científico en lugar de rechazar o aceptar alguna de las posturas cientificistas o dogmáticas, las aborda como hipótesis a contrastar a través de la dialéctica. Por lo tanto, la primera fase consiste en establecer los elementos formales de una tesis y antítesis resultado de las preguntas iniciales a responder, y posteriormente, a través de los elementos de síntesis, establecer las etapas de experimentación, mediación y reflexión. Hasta donde sabemos, el método inspira un abordaje integrador en donde todas las afirmaciones se conviertan en variables medibles y, con ello, promuevan el dinamismo de propiciar conclusiones basadas en procesos claros, lo cual se vuelve indispensable cuando se inicia el estudio de un campo disruptivo o una tecnología emergente.

Tal es el caso de plantear una pregunta inicial formulada en términos cotidianos cargada de ambigüedad y antropomorfismo, como lo es: ¿los LLM mienten? Dicha pregunta puede ser abordada desde enfoques epistemológicos radicales, con fines de evidenciar sus fortalezas y límites dado el planteamiento. La mediación del método científico no busca el abordaje desde un plano moral, sino reformular la intención de la investigación, traduciendo la inquietud original en un problema técnicamente definible, medible y verificable. De este modo, la mediación científica permite sustituir concepciones humanas como la mentira, por categorías operacionales precisas, variables y relaciones causales, así como orientar el diseño experimental, mostrando su ventaja fundamental: transformar una controversia conceptual en una investigación científica con una metodología rigurosa (Figura 1).

Figura 1

Mediación del método científico en la reformulación de una pregunta cotidiana sobre IA y LLM.



Retos de la investigación científica en el siglo XXI

El rol del método científico es pieza clave en la búsqueda de generación de nuevo conocimiento. Una metodología que hemos adoptado desde hace ya varios años con resultados sobresalientes que nos posicionan en un dominio del conocimiento bastante sólido y unificado. No obstante, bajo el lente con el que se examina la realidad actual de la investigación, podemos identificar cuestiones críticas que dificultan un actuar libre y apropiado con el que se pretende establecer el mismo marco robusto del que estamos acostumbrados (denotado nuevamente por las bases del método). Hoy en día, es inevitable no considerar la carrera desenfrenada por "hacer ciencia", marcada por las demandas de innovación que exige la actual academia, además del ritmo vertiginoso del avance tecnológico, entre otras razones. De aquí surgen retos como:

1. Comprender con claridad el enfoque científico que se pretende abordar; basados en parte de los elementos del método (descrito con brevedad en las secciones anteriores de este capítulo), y que radica en la nitidez conceptual para convertirse en un pilar que sostiene la posible complejidad de una investigación.

2. La carrera contrarreloj por publicar material científico de alto impacto (también conocida como *Publish or Perish*), donde el sistema de investigación recompensa la velocidad y visibilidad de los resultados, que son variables no necesariamente vinculadas a un componente de alto impacto.

3. La infraestructura relacionada con los materiales mínimos para efectuar la metodología correctamente; donde incluso las herramientas tecnológicas básicas son primordiales a lo largo del desarrollo de una investigación, con lo que se reduce la credibilidad del trabajo en caso de no contar con una infraestructura idónea.

4. El financiamiento adecuado, vinculado al coste del tiempo de vida de un proyecto de investigación, que suele brindarse en tiempos relativamente cortos, a pesar de que se busca que sus resultados impacten más allá del tiempo de duración del proyecto y del financiamiento. Entre otros.

Podemos identificar un componente que incide directa y ponderadamente en cada uno de los retos antes mencionados, y que mantiene una relación estrecha con la forma actual de hacer investigación: el procesamiento de datos dentro de una era de abundante información. Si bien, la disponibilidad de datos masivos podría parecer una ventaja vinculada a la democratización del conocimiento, también podría generar una paradoja al tener más

información y conocimiento al alcance, pero con menos claridad. Esta situación plantea un conjunto de repercusiones que amplifican los retos ya expuestos, entre las que destacan respectivamente:

1. Empirismo ciego y sobrecarga de información. La tentación es enorme, recolectar un sinfín de datos en la etapa de observabilidad, pero carentes de un objetivo. Más allá de establecer una línea de investigación, pero buscar sin un propósito es hacer ciencia a ciegas, experimentando con recursos mal aprovechados. La magnitud del gran océano actual de información repercute directamente en la manera correcta de procesarlos y plantear evidencia de un fenómeno observable que es establecido con rigor desde los inicios de una investigación.

2. Irreproducibilidad y sesgo masivo en datos masivos. Uno de los objetivos de la investigación científica es hacer reproducible los hallazgos, significa evidenciar que el fenómeno está presente bajo las mismas condiciones y no fue solo un acto de aleatoriedad. Cuando el fenómeno es complejo, posiblemente el análisis también lo sea, por ello la metodología debe describirse con claridad, lo que da lugar a la transparencia. En el caso del sesgo inherente a los datos, simplemente representa un reflejo de la naturaleza humana, ejecutando y amplificando los patrones de comportamiento que repercuten directamente en los sistemas que ocupan esos datos para replicar sus decisiones. Entonces, ¿deberíamos esperar que los sistemas de IA reduzcan el sesgo y provean decisiones humanas, cuando ni siquiera el ser humano tiene la precaución de hacerlo debido a su incompletitud natural?

3. Saturación de información y brecha digital. La abundancia de datos ha transformado la investigación preliminar, ahora se trata de hallar evidencia dentro del mismo proceso de revisiones de literatura y con ello efectuar interpretación significativa. Por lo tanto, la carga de revisar y validar evidencia suele ser abrumadora al manejar grandes volúmenes de datos que requieren un escrutinio meticuloso para considerarse fiables. Esto habla de dejar al descubierto un perfil académico diverso con capacidades de investigación variadas y de metodologías alternativas. La necesidad de desarrollar competencias digitales se vuelve prioridad, además de fomentar la capacitación continua y una cultura de colaboración y transparencia, lo que pone en evidencia una brecha digital marcada dentro de la comunidad científica por sus recursos y formación.

4. Incremento en los costes de financiamiento. Los proyectos de investigación, concebidos para un periodo determinado, suelen enfrentar imprevistos contextuales y

técnicos que extienden su duración. Esta extensión, es particularmente crítica en la fase de procesamiento y análisis de datos, donde efectuar un procesamiento minucioso determinaría incluso la validez de los resultados; con esto, se genera un cuello de botella que incrementa los costos. Esta situación obliga al equipo de investigación a dedicar tiempo crucial en buscar financiamiento adicional, mermando posiblemente el tiempo efectivo invertido a la investigación.

Precisamente, a lo largo del estudio de la disciplina de IA, se han detectado múltiples beneficios, tantos como para posicionarse como una alternativa sólida en la ejecución de tareas mecánicas que no requieren la atención humana como prioridad. Precisamente, el objetivo inicial de esta disciplina dentro de las ciencias de la computación ha sido el mismo desde sus inicios en la conferencia de Dartmouth en 1955: lograr sistemas que ejecuten ciertas tareas al simular habilidades humanas (McCarthy et al., 2006). En efecto, la gran cantidad de elementos publicados en esta disciplina nos orilla a pensar en el ímpetu transformador de la IA como catalizador de progreso. Su impacto ha trascendido al grado de convertirse en un hito tecnológico (a la altura de internet). Entonces, ¿sabe algo la investigación científica que ha dedicado muchos esfuerzos en estudiar la IA? Quizás solo sea una moda o probablemente sea un instinto propio de querer renovarse. ¿No nos ha enseñado la investigación científica a prestar atención a las señales de nuestro alrededor?

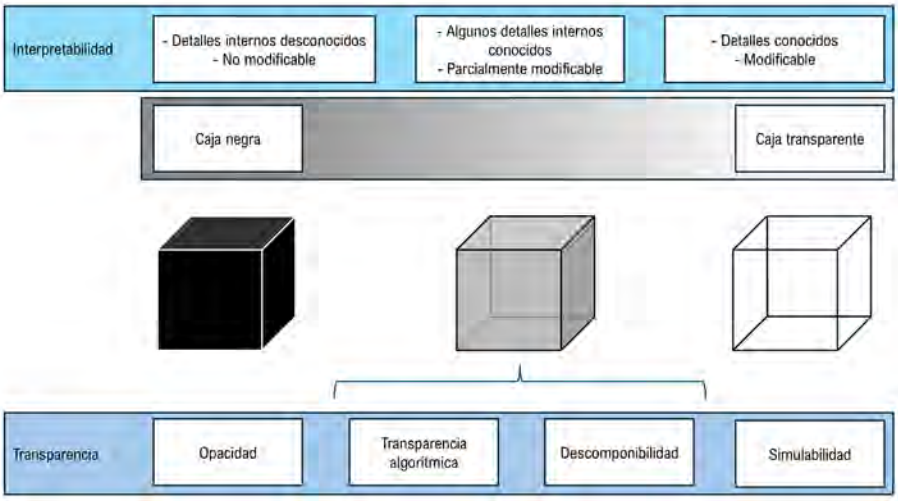
Una mirada ética desde las ciencias de la computación

Se ha mencionado la importancia de la transparencia al llevar a cabo investigación, la transparencia no solo revela el proceso metodológico, sus hallazgos y carencias, sino que también inspira confianza, certeza y rigor (McCormack et al., 2025). De forma muy similar, en ciencias de la computación, se dice que un sistema transparente es aquel que revela su proceso interno de condiciones y variables necesarias y suficientes para ejecutar una acción, con lo que simplifica su comportamiento haciéndolo, en teoría, más entendible (como una caja transparente) (Castelvecchi, 2016). De forma opuesta, hallamos una relación entre el concepto de cajas negras, con los elementos de ambigüedad, incertidumbre y opacidad. En efecto, el término se refiere a la falta de comprensión de cómo un sistema transforma una entrada definida a una salida que incluso podría ser una decisión. El enfoque de ciencias de la computación pretende abordar el estudio de la transparencia identificando el nivel de interpretabilidad de un sistema. Lo que implica responder preguntas como: ¿qué tanto sabemos de un sistema que ejecuta acciones para comprender sus decisiones?

La interpretabilidad sirve de puente entre la comprensión de un sistema y su ejecución, con claras ventajas si lo quisiéramos aplicar a una tarea específica, puesto que tales acciones caerían dentro de nuestra comprensión por la manera de conocer el camino que condujo a tomarlas. Ahora bien, más allá de posicionar la necesidad de establecer una clasificación entre dos elementos opuestos como son cajas transparentes y cajas negras, un objetivo del campo de la explicabilidad es conocer qué sistemas cumplen con determinadas condiciones para posiblemente ser situados en un punto intermedio y con ello, abordar la manera en cómo convertir algún proceso interno en transparente (Figura 2). Es decir, la explicabilidad estudia precisamente las técnicas y métodos adecuados para establecer un criterio entre lo claro y lo opaco. Con ello se pretende transmitir el conocimiento apropiadamente entre humanos y máquinas de manera simbiótica.

Figura 2

De cajas negras a transparentes, sus definiciones en Interpretabilidad, Transparencia y la promovida por Ciencias de la Computación.



No resulta trivial el diseño de sistemas comprensibles, simples y precisos. De hecho, la paradoja de la complejidad resalta que muchos de los modelos más potentes son inherentemente menos interpretables (Rudin, 2019). La explicabilidad no representa un atributo innato de sistemas de IA, ya que dentro de los principales objetivos de la disciplina de IA se basa en buscar la precisión independientemente del coste computacional; es decir, les hemos diseñado para optimizar los procesos que buscan una solución que nos sea de utilidad. Así pues, el dilema surge al preguntarnos si ¿deberíamos sacrificar rendimiento por explicabilidad?

La transparencia y la interpretabilidad constituyen pilares esenciales en el desarrollo y uso de sistemas de IA. Sin embargo, con frecuencia pasan a segundo plano frente a otros principios que han adquirido mayor visibilidad o resonancia mediática, como la seguridad, la justicia, la responsabilidad, la gobernanza, la contestabilidad o la reparación (GOV.UK, 2023; OECD, 2024). Tanto la transparencia como la explicabilidad han construido un espacio importante en el estudio de las ciencias de la computación, y proveen la perspectiva que buscamos en otras disciplinas. Un gran hito de innovación tendrá lugar al momento de transformar conceptos técnicos a éticos y viceversa, y hacer que cada componente de IA refleje los principios fundamentales guiados por normas humanas mediante un lenguaje de máquina apropiado.

El hito de la inteligencia artificial generativa

A lo largo de más de 70 años, la disciplina de la IA se ha transformado al desarrollar diferentes enfoques y herramientas, a la par de que los especialistas del área buscan englobar todas ellas mediante una clasificación entendible por todos. Por ejemplo, la manera de entrenar un sistema dependerá de la forma en cómo le mostramos los datos (que muy probablemente contengan algún tipo de sesgo humano). Esta metodología de entrenamiento de máquina pareciera similar a las teorías del aprendizaje humano, lo que sugiere algún tipo de relación entre humanos y máquinas que nos orilla a creer que no estamos tan alejados. Toda esta teoría da lugar al gran campo del aprendizaje de máquina (mejor conocido como *Machine Learning*) que incide en muchas de las plataformas digitales que conocemos ahora y que en ocasiones pasa desapercibido debido a su uso rutinario. De aquí se desprenden otras formas de ver a los sistemas de IA dependiendo de cuál sea su tarea predilecta u objetivo de diseño, ya sea clasificar o generar.

Cuando un sistema tiene la capacidad de clasificar, se dice que pertenece al grupo de IA discriminativa. Estos sistemas tienen la facultad de tomar una decisión sobre los datos que procesan y asignarles una etiqueta particular. Suena sencillo, aunque su diseño conlleva ejecutar alguno de los paradigmas de entrenamiento de máquina, por lo que su implementación no es tan trivial. De manera similar, aunque en cuestión de funcionamiento, un sistema que tiene la capacidad de procesar la información que se le brinda y tomar una decisión con fines de crear un nuevo dato (dato artificial), se dice que pertenece al grupo de IAGen. Al parecer, este último grupo ha sido el de mayor relevancia en los últimos años y ha tomado la batuta para producir nueva información. Se discute en muchos espacios sobre los beneficios de la IAGen; por ejemplo, la redundancia de información que incluso

los mismos sistemas de IA requieren para mejorar sus resultados. La ecuación es sencilla de plantear: a medida que los datos masivos crecen, estos constituyen un elemento primordial para mejorar el desempeño de los modelos de IA, aunque se trate de datos artificiales. Por otro lado, como hemos visto en pasajes anteriores de este texto, datos masivos en contextos humanos repercuten en la facilidad de procesamiento y, por lo tanto, en nuestro entendimiento.

Las consecuencias son diversas, no obstante, todas guardan relación sobre el actual reto de distinguir qué información es creada por humanos y cuál es la creada artificialmente. ¿Es acaso que alguna pierde validez? ¿A cuál debemos creerle? La IAGen representa un hito precisamente por los planteamientos que surgen a medida que la acoplamos a nuestras vidas. Conceptos que anteriormente pensábamos tener dominados con holgura regresan a nuestras mentes con duda. Es ahí el mayor aporte de la IAGen en nuestra era, más que continuar el discurso clásico actual de enunciar beneficios y limitaciones, es el ejercicio de introspección que ha traído consigo para replantear diversos elementos de las ciencias, las humanidades, la investigación e incluso nuestra trascendencia en el universo.

Elementos de un nuevo método

Se sabe que el método científico ha cambiado paulatinamente, adoptando matices apoyados en los paradigmas de cada época. Hoy somos conscientes que los sesgos humanos impactan no solo en los datos, sino que prevalecen en la propia investigación. En el contexto del método científico reconfigurado por los datos masivos y la inteligencia artificial generativa, puede plantearse la hipótesis de que toda producción de conocimiento está atravesada por dos formas distinguibles de distorsión: el sesgo epistemológico y el sesgo aleatorio. El primero consiste en una desviación sistemática que surge de los marcos conceptuales, decisiones metodológicas, arquitecturas instrumentales y modelos interpretativos desde los cuales se construye la observación; es decir, no proviene del fenómeno estudiado, sino de la estructura cognitiva y técnica que lo interroga. Es acumulativo, reproducible y potencialmente corregible mediante mecanismos de transparencia, trazabilidad y pensamiento crítico.

El segundo, en contraste, corresponde a la variabilidad estocástica inherente al propio sistema observado: fluctuaciones estadísticas, ruido experimental o indeterminaciones físicas que no pueden eliminarse completamente, sino solo modelarse probabilísticamente. Mientras el sesgo epistemológico delimita lo que somos capaces de ver y cómo lo interpretamos,

el sesgo aleatorio marca el límite ontológico de lo que puede predecirse con certeza. La solidez del nuevo método científico dependería, entonces, de su capacidad para distinguir ambos planos: mitigar estructuralmente el primero e integrar formalmente el segundo dentro de sus modelos explicativos.

Podríamos incluso estudiar la forma de mitigar los sesgos si identificamos sus fuentes y adoptamos una postura en la que no se confíe en la objetividad del científico, sino que diseñe un marco para confiar en la objetividad del método. De hecho, el falsacionismo de Popper es una medida profunda que permite saber cómo nuestros sesgos distorsionan la búsqueda y generación del conocimiento. Se afirma que incluso una teoría no puede ser demostrada definitivamente como verdadera, sino que solo se corrobora mientras resiste intentos rigurosos de falsarla o refutarla (Popper, 1934). Cada componente del método parece trascender su forma original para acoplarse a un paradigma más robusto, dinámico y flexible. Esto le permite ser apto para abordar los desafíos emergentes de la ciencia moderna, entre los que destacan:

- Colaboración interdisciplinaria. Pasa de ser una opción a una necesidad. Las fronteras tradicionales entre disciplinas se desvanecen rápidamente debido a que el estudio de cada fenómeno particular repercute en otros campos por su complejidad y escala. Los enfoques interdisciplinarios buscan facilitar el sentido de innovación y mejorar la calidad de la ciencia. Representa el sueño académico, lograr un equipo de trabajo con diferentes perfiles.
- Marco ético transversal. Implica considerar aspectos técnicos y metodológicos alineados con las directrices de un nuevo método basado en principios. Hasta hace algunos años, las implicaciones de una investigación se ejecutaban a-posteriori, es decir, en las etapas finales del método. Sin embargo, con el impulso de un enfoque ético transversal y la utilidad de la trazabilidad de las decisiones desde etapas tempranas, se fomenta un mayor sentido de transparencia metódica. Concepto similar al definido por las ciencias de la computación como transparencia, con el que se pretende asegurar la integridad y reproducibilidad de sus experimentos y análisis.
- Ciencia guiada por datos. Inevitablemente surgirán dudas justo después de maniobrar con datos recopilados por algún proceso de observación. Dudas que darán lugar al descubrimiento de nuevos patrones y con ello planteamientos novedosos de investigación. Parecería que los papeles se invierten, de hipótesis-análisis a

análisis-hipótesis al asegurarnos primero si existe evidencia en los datos que no habíamos considerado. Sin embargo, podemos contemplar a la ciencia guiada por datos, como un elemento transformador del proceso de observabilidad y le robustece. De hecho, los sistemas de IAGen son muy buenos hallando patrones (les hemos entrenado para ello por años), lo que ocasiona llevar el razonamiento inductivo a grandes escalas, tal como hemos presenciado a raíz de la consolidación de los datos masivos y LLM.

- Impulso de la ciencia abierta. Una era de ciencia abierta está marcada por promover el acceso libre a la información y recursos científicos, permitiendo que los grupos de investigación mejoren los conocimientos obtenidos y aceleren el ritmo de innovación. Este enfoque pretende abrir la puerta a la colaboración global y construir una red de conocimientos sostenible e interrelacionada.

Reflexiones finales y conclusiones

En un mundo donde las ciencias de la computación aportan en demasía a diversas disciplinas, el siglo XXI ha visto a este campo como motor transformador, ofreciendo no solo conocimientos teóricos y técnicos, sino también elementos de análisis y síntesis. Dentro de estas contribuciones, la IA destaca por su papel en ramificaciones cada vez más avanzadas y complejas de la misma, como la IAGen, que ha sido considerada un hito no solo en el ámbito académico, científico o tecnológico, sino también en la forma en que percibimos y actuamos en la vida cotidiana. La IAGen no se limita a ser un mero objeto de estudio o una herramienta para otras ciencias; ha venido a desafiarnos al replantear conceptos y elementos fundamentales como la conciencia, la inteligencia, el aprendizaje y la docencia. Ha generado consigo reflexiones profundas que nos obligan a reformular nuestra comprensión de diversos temas y ofrecer maneras alternativas de llevar a cabo ciencia. Es precisamente lo que la hace disruptiva.

En el campo de la investigación científica, parece haber un indicio de cambio paradigmático. La justificación se basa en los elementos transversales innovadores que reinventan enfoques clásicos y superan retos emergentes. La interdisciplinariedad, los marcos éticos, la transferencia tecnológica, la ciencia abierta y guiada por datos, entre otros, se convierten en necesidades actuales que buscan impactar en mayor escala a la innovación y desarrollo. Como en lo referente a las propuestas actuales en el desarrollo de proyectos de IA, donde se debería tener como objetivo principal el abordar el gran reto de consolidar la teoría de

un cambio de paradigma a la vez que intenta transcribir nuestros pensamientos y necesidades en lenguaje técnico, y apoyar la misma investigación científica en su búsqueda por renovarse y reinventarse.

Referencias

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23. https://doi.org/10.1038/538020a

Descartes, R. (1637). *Discours de la méthode*. Jan Maire.

GOV.UK. (2023, agosto 3). A pro-innovation approach to AI regulation. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#annexc

Hume, D. (1748). Philosophical essays concerning human understanding.

Johnson, A. (2023, enero 31). *ChatGPT In Schools: Here's Where It's Banned—And How It Could Potentially Help Students*. Forbes. https://www.forbes.com/sites/arianajohnson/2023/01/18/chatgpt-in-schools-heres-where-its-banned-and-how-it-could-potentially-help-students/

LaPlace, P. S. (1814). *Essai philosophique sur les probabilités*. M.me V.e Courcier.

Lorenz, E. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Science*, *20*, 130-141.

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence: August 31, 1955. *AI magazine*, *27*(4), 12–14. https://doi.org/10.1609/aimag.v27i4.1904

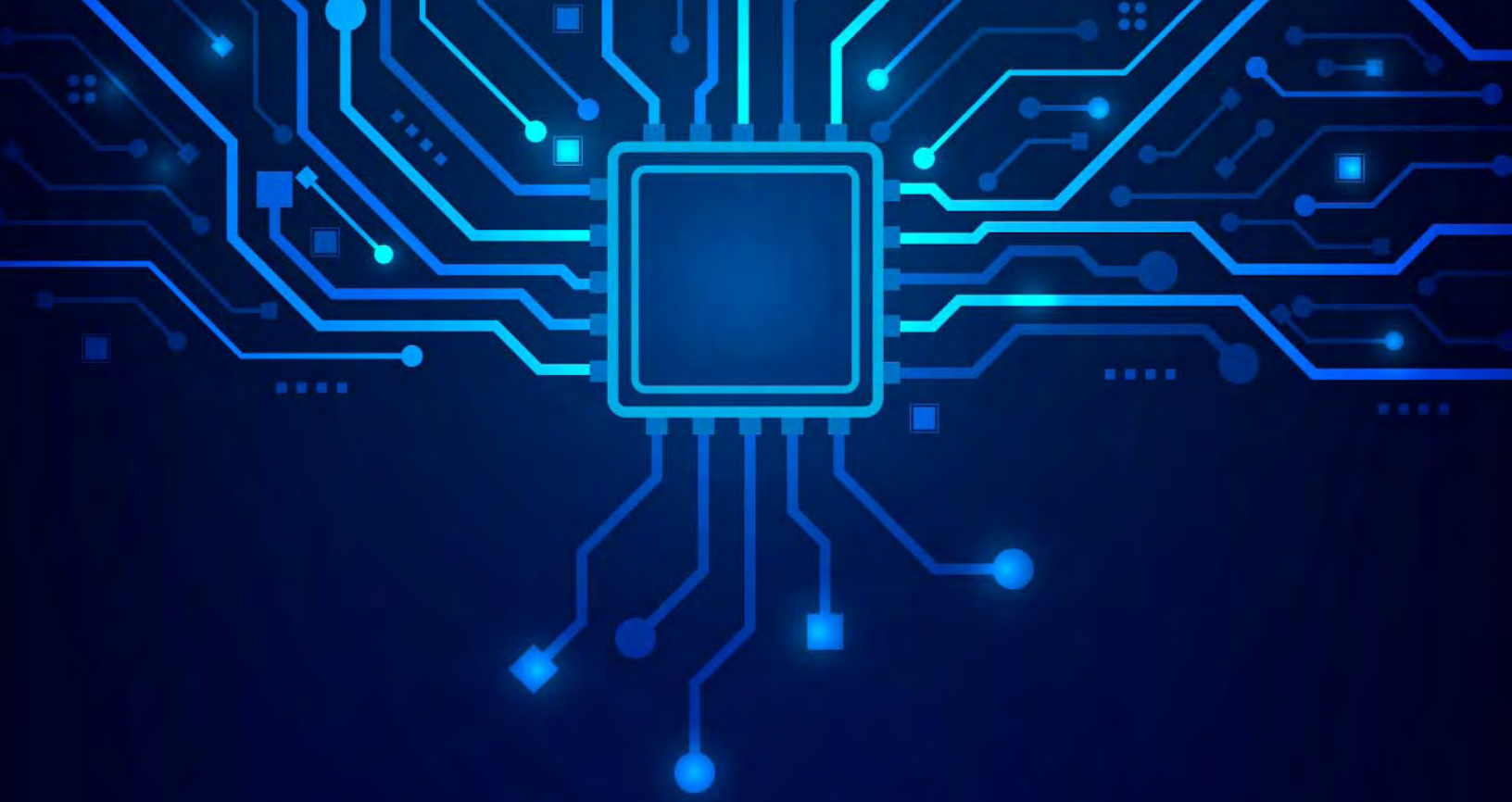
McCormack, L., Bendechache, M., Lewis, D., & Huyskes, D. (2025). Trust and transparency in AI: Industry voices on data, ethics, and compliance. *AI & SOCIETY*. https://doi.org/10.1007/s00146-025-02654-7

OECD. (2024, mayo 1). *OECD AI Principles overview*. https://oecd.ai/en/ai-principles

Popper, K. (1934). *Logik der Forschung*. Julius Springer.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sutton, R. S., & Barto, A. (2020). *Reinforcement learning: An introduction* (2nd ed.). The MIT Press.



12.

La inteligencia artificial en el derecho: uso práctico y transición regulatoria

Carlos Enrique Levet Rivera
Universidad Veracruzana, México
ORCID: 0009-0002-8784-0549

Modesta Lorena Hernández Sánchez
Universidad Veracruzana, México
ORCID: 0009-0006-8686-6412

Ramar Mendoza Díaz
Universidad Veracruzana, México
ORCID: 0000-0003-2836-2364

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.12

12.

La inteligencia artificial en el derecho: uso práctico y transición regulatoria

Introducción

La inteligencia artificial (IA) se ha consolidado como uno de los avances tecnológicos más significativos del siglo XXI, debido a su capacidad para transformar de manera profunda los procesos sociales, económicos, educativos y jurídicos. Su desarrollo ha permitido la creación de sistemas capaces de analizar grandes volúmenes de información, aprender patrones históricos, automatizar tareas complejas y generar respuestas que antes dependían exclusivamente del razonamiento humano. Este escenario ha generado un impacto directo en el ámbito del derecho, disciplina que históricamente se ha venido adaptando a los cambios derivados del progreso científico y tecnológico, enfrentándose al desafío de la integración de herramientas algorítmicas en la función jurisdiccional, sin comprometer los principios fundamentales que sostienen la justicia y los pilares esenciales del Estado de Derecho.

Lo anterior confirma una realidad conocida por todos: la incorporación de la IA en el ámbito jurídico ha abierto nuevas posibilidades para la optimización de procesos judiciales y administrativos, tales como, el análisis de jurisprudencia, la gestión documental, la predicción de riesgos procesales, la evaluación de evidencias y el apoyo en la toma de decisiones jurisdiccionales. Estas aplicaciones si bien es cierto que, prometen una mayor eficiencia, rapidez y coherencia en la administración de justicia, no menos cierto es que, también han suscitado un intenso debate respecto de los riesgos que implica delegar funciones decisorias a sistemas automatizados, especialmente cuando se encuentra en disputa la resolución de controversias que envuelven garantías fundamentales como el principio de legalidad, el debido proceso, la presunción de inocencia, la motivación de las resoluciones y el derecho a una defensa adecuada.

Asimismo, se explora el papel de la IA como apoyo en la enseñanza y el ejercicio profesional del derecho, destacando su potencial para fortalecer el aprendizaje activo, la formación de habilidades analíticas, la interpretación normativa y la argumentación jurídica. Esto implica una aplicación ética, crítica y responsable, evitando una dependencia a dicha tecnología.

En ese orden de ideas, la presente colaboración tiene como finalidad analizar desde una perspectiva jurídica general, los principales usos de la IA en las actividades legales, así como su regulación progresiva en el mundo. Durante su desarrollo se realizan planteamientos doctrinales, jurisprudenciales y normativos, que coadyuvan a brindar una postura crítica sobre su implementación en la práctica jurídica contemporánea, así como en los desafíos que enfrenta la delimitación de sus funciones en la legislación vigente.

Intersección de la inteligencia artificial en el ámbito jurídico

Aunque hoy en día se le denomina IA a cualquier avance tecnológico que simula un comportamiento humano -*Chat GPT*, *Siri*, *Alexa*, etc.-, lo cierto es que, se refiere a la evolución que ha tenido la informática para ejecutar procesos con rapidez en diversas áreas, como: salud, industria, educación, y desde luego, el sector legal.

En esencia, la IA tiene como propósito que los ordenadores produzcan la misma clase de cosas que puede crear la mente; por ello, su capacidad para razonar la describe como *inteligente*. Sin embargo, el litigio y la impartición de justicia necesariamente requieren de otro tipo de aptitudes que difícilmente puede imitar una máquina inteligente a través de la repetición, simulación y/o predicción; pero es una realidad que el ámbito legal se encuentra transitando hacia un modelo en el que la sombra de los algoritmos se perfila para dirigir el destino de las decisiones jurídicas. Lo anterior, ha propiciado que el comportamiento del razonamiento automatizado, se analice y discuta en tareas específicas como las siguientes.

Psicoterapia e inteligencia artificial

Actualmente, las técnicas ericksonianas –provenientes del campo de la psicoterapia-, son un modelo idóneo para extraer y estructurar el conocimiento judicial implícito de los conocedores del derecho. Dichos procesos fueron desarrollados por Milton Erickson (Haley, 1997), y representan una oportunidad viable que puede adaptarse a la tecnología que funciona con IA, para mejorar los métodos de razonamiento judicial y toma de decisiones.

En la opinión de Enrique Cáceres Nieto (2016), los jueces deciden con base en procesos y esquemas generados por el inconsciente adaptativo, dentro de un contexto de aprendizaje implícito. Por lo que, bajo esos criterios, los sistemas expertos tienen la posibilidad de alimentarse del conocimiento que pueda extraerse a través estas técnicas, para replicar patrones de razonamiento judicial.

En ese orden de ideas, las técnicas eriksonianas pueden lograr que los jueces accedan a las estructuras de su propio consciente adaptativo, es decir, aquella parte de su mente que actúa de forma automática y rápida, tomando decisiones o realizando tareas, sin la necesidad de pensar conscientemente en el proceso que se lleva a cabo para ejecutarse. Esto ayuda a adaptarse al entorno y responder de manera intuitiva de acuerdo a experiencias pasadas.

De este modo, se plantea que las máquinas inteligentes se programen bajo un enfoque de exploración y explicación del consciente adaptativo, con la finalidad de que los jueces tomen decisiones expeditas y eficientes en resoluciones complejas, basándose en el conocimiento acumulado, con alcance para reconocer patrones en casos legales análogos, sin la necesidad de analizar minuciosamente cada detalle. Asimismo, se proyecta que el juez pueda actuar diligentemente en situaciones urgentes o bajo presión, tomando decisiones instintivas perfectibles con un análisis consciente posterior; esto conduce a enfocarse en los aspectos más importantes de una controversia, reduciendo la sobrecarga cognitiva.

Avances de lo antes expuesto, se han materializado en la máquina denominada *Expertius I* (Cáceres Nieto, 2023), que responde a un modelo mental basado en el constructivismo jurídico de acuerdo al cual los jueces resuelven las controversias a través de la sistematización cognitiva de diferentes clases de elementos –derecho positivo, esquemas doctrinales, jurisprudencias, información relevante como las afirmaciones hechas por el demandante y el demandado, así como material probatorio-. Su diseño se originó en el Instituto de Investigaciones Jurídicas y el Centro de Ciencia Aplicada y Desarrollo Tecnológico de la Universidad Nacional Autónoma de México, bajo el auspicio del Consejo Nacional de Ciencia y Tecnología, en colaboración del Tribunal Superior de Justicia del Estado de Tabasco, con el objetivo de coadyuvar en la toma de decisiones y estandarizar el conocimiento colectivo de la comunidad judicial aplicable en el juicio especial de alimentos. El espécimen está dirigido a brindar asesoría a los jueces y secretarios mexicanos novatos, para determinar si el demandante es elegible o no para otorgarle una pensión alimenticia y en su caso, decretar el monto correspondiente.

Por otro lado, México también cuenta con un desarrollo más sofisticado denominado *Expertius II*, el cual fue desarrollado en el Laboratorio de Constructivismo Jurídico del Instituto de Investigaciones Jurídicas de la Universidad Nacional Autónoma de México, bajo el auspicio del programa Fronteras de la Ciencia del Consejo Nacional de Ciencia y Tecnología (Cáceres Nieto, 2023). Funciona con una epistemología jurídica aplicada enfocada en la definición de los procesos cognitivos que deben ejecutar los operadores jurídicos para determinar la

verdad de los hechos jurídicamente relevantes. Su objetivo consiste en aportar las razones jurídicas idóneas para justificar una decisión judicial, con la intervención de algoritmos que sean capaces de identificar patrones y sean compatibles con la transparencia bajo la cual la máquina resuelve.

Implementación algorítmica de la regla de la lenidad

Se ha dicho que la interpretación jurídica casi siempre admite diversas opciones loables, siempre y cuando estén fundamentadas racionalmente en resultados hermenéuticos posibles; en ese sentido, bajo el criterio de *más probable que no*, el algoritmo puede aplicar un parámetro estadístico entre los precedentes jurisprudenciales, seleccionando aquel que tenga mayor probabilidad de ser cierto, por ser de uso frecuente. Tal es el caso de *PROMETEA*, un *software* que empezó a utilizarse en el 2017 por la Fiscalía General Adjunta en lo Contencioso Administrativo y Tributario del Ministerio Público Fiscal de Buenos Aires, para resolver casos, prediciendo situaciones y hechos sobre los que existe una tendencia palpable en criterios jurisprudenciales (Estevez et al., 2020). Otro ejemplo es *PretorIA* (2020), implementada por la Corte Constitucional de Colombia en el año 2020, con funciones específicas de búsqueda, análisis y clasificación de información de las sentencias, categorización y generación de estadísticas. Cabe destacar que no realiza predicciones, ni proyecta decisiones –a diferencia de *PROMETEA*- porque su enfoque consiste en definir el universo de escogencia.

En virtud de lo anterior, queda claro que la IA proporciona herramientas basadas en un enfoque estadístico, con capacidades avanzadas de cálculo y autoaprendizajes, las cuales están preparadas para captar las similitudes que existen en la información y proporcionar rápidamente una respuesta. No obstante, se considera que, al permitir que el algoritmo basado en precedentes judiciales opere de acuerdo a la regla *más probable que no*, generaría mayor incertidumbre sobre todo en controversias del derecho penal, precisamente porque la regla de la lenidad tiene como finalidad proteger los derechos humanos y beneficiar a una persona imputada, frente a los vacíos o imprecisiones que presente la norma jurídica, y/o en su caso, frente a las operaciones estadísticas de la IA. Aunque se asegure que la máquina inteligente trabaja en pro de la tutela judicial efectiva, garantizando la protección de los derechos humanos de las partes involucradas, existe la incertidumbre sobre comportamientos que distorsionen la realidad y, por ende, repercuta en la decisión final que se dicte en una controversia.

Predicción de riesgo procesal

Una de las aplicaciones más debatidas, es el uso de algoritmos para evaluar el riesgo de fuga, reincidencia o peligro para la víctima. Estos sistemas pueden apoyar en la toma de las decisiones relacionadas con la implementación de medidas cautelares, como la prisión preventiva o la imposición de medidas alternativas, contribuyendo a disposiciones más proporcionales y fundamentadas, siempre bajo control judicial. Ejemplo de esta función se materializa en el programa *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS, por sus siglas en inglés), diseñado en Estados Unidos para evaluar a partir de un cuestionario, las posibilidades que tiene una persona encarcelada en reincidir en la comisión de un delito (Salas, 1998). No obstante, este programa inteligente ha dado margen a situaciones violatorias de derechos fundamentales y cuestionamientos vinculados a su falta de transparencia, como el caso *State of Wisconsin vs. Loomis*, promovido en el año 2015 ante la Suprema Corte del Estado de Wisconsin, Estados Unidos, teniendo como hecho controversial errores de sesgo racial en perjuicio del ciudadano Eric L. Loomis, al haberle decretado una condena de seis años de prisión y cinco de condicional, poniendo en entredicho su derecho de presunción de inocencia (State of Wisconsin *vs*. Eric L. Loomis, 2015) y confirmando que los jueces estuvieron aplicando los criterios de *COMPAS* en sus sentencias, sin cuestionar el posicionamiento de la IA.

De igual modo, existen algoritmos que fungen de apoyo en las decisiones jurídicas, dado a su capacidad de analizar jurisprudencia, criterios doctrinales y resoluciones previas, para ofrecer escenarios comparativos en casos similares. Al respecto, *Ross Intelligence* (2016), es un representante legal virtual diseñado en Toronto, Canadá, que a partir del año 2016 funciona en la búsqueda de fundamentos legales, criterios jurisprudenciales, lecturas rápidas de expedientes y sentencias, estructuración de argumentos y juicios de valor, a fin de ofrecerle al operador jurídico, así como a las partes involucradas en una controversia, la oportunidad de preparar una defensa adecuada, brindar mayor coherencia en las decisiones judiciales y reforzar la seguridad jurídica, sin comprometer la independencia del juez ni el principio de libre valoración de la prueba.

Confiabilidad

En cualquier proceso judicial se manejan grandes cantidades de datos, como pruebas documentales, peritaje, registros telefónicos y antecedentes judiciales, que los sistemas de IA pueden procesar y analizar de manera rápida y precisa, facilitando la detención de

patrones relevantes, inconsistencias o conexiones entre los hechos, que podrían pasar desapercibidos en un análisis realizado por una persona. Por ello, estos programas inteligentes se han visibilizado como alternativas para minimizar los errores derivados del cansancio, la sobrecarga de trabajo o la omisión voluntaria de información relevante; sin embargo, la divergencia entre el componente antropológico del derecho y su renovada apariencia algorítmica, abre el camino a dilemas éticos y jurídicos vinculados con la ausencia de una normativa específica en el tema, que inevitablemente repercuten en el papel de la lógica decisoria de la interpretación de la ley.

Es cierto que la objetividad del juez autómata, que esperaba Montesquieu, contrasta con la subjetividad del juez emocional moderno; situación que permite advertir la tensión que existe entre el deseo de mecanizar el proceso y los riesgos vinculados a la arbitrariedad de la subjetividad. Por ello, se puede definir que el dúo *juez y algoritmo*, representan la cara de la misma moneda, porque a ambos les interesa la impartición de la justicia; de ahí que, resulte necesario encontrar un equilibrio entre la justicia digital y la humana; lo cual no es una tarea fácil, si se considera que la justicia impulsada por la tecnología, no se basa al cien por ciento en la ley, sino en un modelo algorítmico cimentado en precedentes jurisprudenciales; a diferencia de la justicia humana, que permite la libertad de juicio, la incertidumbre del resultado y la discreción del juez.

En ese sentido, María Princi (2024) sostiene que es importante considerar la opacidad del algoritmo que estructuralmente resulta incompatible con la carga de la motivación, y conduce inevitablemente a un choque en los derechos humanos y garantías constitucionales; de modo que, los beneficios potenciales de una decisión algorítmica son aceptables porque transforma la información, pero también es cierto que existe incertidumbre de que distorsione la realidad y, en consecuencia, violente derechos humanos; esto refuerza el valor que tiene el componente humano y emocional de un juez, en el análisis de los hechos del caso, la interpretación de las normas y la resolución de una controversia, filtrándose en todos los aspectos jurídicos.

Todo ello conduce a que la confiabilidad de los sistemas inteligentes constantemente se vea cuestionada, ya sea por su forma opaca de operar –*caja negra* algorítmica- que impide que las personas afectadas comprendan la forma en la que han sido sancionadas; y, en consecuencia, vulnera el derecho al debido proceso, dado a que la falta de exhaustividad, imposibilita a cualquier persona a ejercer su derecho de defensa. Otro punto debatible, son los prejuicios o sesgos que pueden surgir debido a los datos registrados en la máquina,

basados en cuestiones de género, clase, etnia u otras características sociales. En la misma línea de argumentación, Padilla Sanabria (2025), alude que uno de los efectos negativos de confiar excesivamente en los sistemas automatizados, es la posible reducción o eliminación de la supervisión humana crítica, elemento indispensable en cualquier decisión que transgreda derechos; puntualizando que, dejar la resolución de una controversia al arbitrio de un programa inteligente que carece de la ponderación humana, fortalece la brecha de acceso a la justicia y coloca en una posición de desventaja a las personas afectadas, frente a las autoridades.

La inteligencia artificial en la enseñanza del derecho

Históricamente, la enseñanza del derecho ha estado basada en métodos tradicionales centrados en la exposición magistral del docente, el estudio de textos normativos y doctrinales, y la memorización de conceptos jurídicos. Si bien estos métodos han sido fundamentales para la formación jurídica, evidentemente presentan limitaciones frente a las demandas actuales del ejercicio profesional. Esto ha demostrado que resulta insuficiente formar juristas que solo conozcan las normas, sin capacitarlos para analizarlas, interpretarlas y aplicarlas en contextos complejos y cambiantes.

Ante este panorama, la incorporación de la IA en la enseñanza del derecho se presenta como una alternativa para innovar en los procesos educativos, fomentar el aprendizaje activo y preparar a los estudiantes –futuros profesionistas- para los desafíos exponenciales del siglo XXI.

Actualmente, la IA como herramienta pedagógica, ofrece múltiples posibilidades para enriquecer la educación jurídica, puesto que su implementación no busca reemplazar al docente ni al razonamiento jurídico humano, sino complementar y fortalecer los procesos de enseñanza y aprendizaje.

Uno de los principales aportes de la IA en la educación es la posibilidad de personalizar el aprendizaje. A través del análisis de datos sobre el desempeño académico de los estudiantes, los sistemas de IA pueden identificar patrones, detectar dificultades y adaptar los contenidos a las necesidades individuales. Específicamente en la enseñanza del derecho, ofrece material concreto para reforzar áreas como: la interpretación normativa, la argumentación jurídica o el análisis de casos. En el mismo sentido, permite el desarrollo de simuladores jurídicos que recrean escenarios reales de litigio, negociación o mediación,

lo cual facilita el aprendizaje practico y permite a los estudiantes experimentar la toma de decisiones jurídicas en un entorno controlado.

En el mismo sentido, el uso de sistemas de IA para analizar grandes volúmenes de jurisprudencia y legislación, constituye una herramienta pedagógica de gran valor, dado a que los estudiantes pueden identificar tendencias jurisprudenciales, construir líneas de precedentes y comprender la evolución de criterios judiciales.

En este escenario, el profesional del Derecho tiene la responsabilidad de formarse para comprender los alcances y las implicaciones de la IA, asumiendo un rol activo en su regulación y aplicación. Pues si bien es cierto, la automatización de tareas repetitivas y la estandarización de propuestas de resolución de conflictos, pueden optimizar recursos, pero requieren una atención prioritaria a la transparencia de los algoritmos, la protección de la privacidad y la fiabilidad de los sistemas (Quirós Fons, 2025).

Avances en las disposiciones normativas de la inteligencia artificial

Italia

El 10 de octubre de 2025 entró en vigor la Ley número 132/2025 denominada *Disposizioni e deleghe al Governo in materia di intelligenza artificiale* que regula el uso transparente y responsable de sistemas inteligentes (Cámara de Diputados y Senado de la República de Italia, 2025). A grandes rasgos, las disposiciones italianas regulan aspectos relacionados con el derecho penal: uso de IA en la comisión de delitos, generación o alteración de contenidos a través de IA (*deepfakes*) y manipulación de mercado; en el mismo orden de ideas, establece una gobernanza institucional bajo un modelo de distribución de funciones –omite crear una autoridad específica-; es decir, opta por asignarle competencias a ciertos organismos existentes, los cuales operan a través de la coordinación de un Comité Interministerial; igualmente, contempla los principios relacionados con la legalidad, equidad, respeto por la dignidad humana, transparencia, no discriminación y respeto de los derechos fundamentales; ratificando que en todas aquellas decisiones automatizadas que representen una afectación a los derechos fundamentales de las personas, necesariamente deberá intervenir el raciocinio humano.

Al mismo tiempo, en aras de preservar la transparencia y evaluar el impacto ético y social, la normativa italiana impone la obligación de auditar e implementar mecanismos de verificación de la información, para las empresas, instituciones y profesionales que ejecuten actividades con el apoyo de algún sistema inteligente.

No obstante lo anterior, en dicho texto legal se advierte una omisión vinculada con el uso de la IA en la función jurisdiccional, dado a que, no puntualiza sobre las hipótesis en las que el juez se apoye de programas inteligentes para elaborar las sentencias, así como la forma en la que debe validarse ese contenido. En el mismo tenor, deja un vacío legal en las situaciones en las que el raciocinio humano sea sustituido por un algoritmo, en perjuicio de las partes que intervienen en una controversia; puesto que, no aborda abiertamente el riesgo de sesgos en las decisiones tomadas con intervención de IA.

Independientemente de los puntos a favor o en contra que se pueden advertir en la ley italiana, lo cierto es que los profesionales del derecho están obligados a actualizarse en las materias sobre derecho digital, protección de datos personales, responsabilidades derivadas del uso de las tecnologías, entre otros; a fin de transitar adecuadamente en dicho ecosistema tecnológico.

España

El 30 de enero de 2026 se publicó la instrucción 2/2026 del 28 de enero, del Pleno del Consejo General del Poder Judicial, sobre la utilización de sistemas de IA en el ejercicio de la actividad jurisdiccional (Pleno del Consejo General del Poder Judicial de España, 2026). Esencialmente, la normativa española contempla que el juez sí puede utilizar programas inteligentes en la función jurisdiccional, siempre y cuando, éstos sean facilitados por el Poder Judicial de aquel país, -excluye cualquier tipo de programa de IA comercial-; delimita la función de la IA exclusivamente para brindar asistencia en la búsqueda y clasificación de información y/o actuaciones, elaboración de esquemas y resúmenes, así como apoyo en tareas organizativas, precisando que las decisiones automatizadas no pueden sustituir el juicio valorativo e interpretación que emita el juez.

Aunado a lo anterior, también se establece que el juez puede utilizar la IA para elaborar borradores de sentencias; con la salvedad de que, dicho resultado debe ser revisado de forma personal y crítica por el operador jurídico. Dicho proceder debe ajustarse al principio de control humano efectivo, principio de no sustitución de los jueces, principio

de responsabilidad judicial, principio de independencia judicial, principio de respeto a los derechos fundamentales, principio de confidencialidad y seguridad, principio de prevención de sesgos algorítmicos, principio de proporcionalidad y uso limitado. El incumplimiento de la normativa en comento, puede ocasionar responsabilidades conforme a la Ley Orgánica del Poder Judicial de España.

México

Actualmente no existe una ley específica y vigente que regule de manera integral el uso de la IA en el derecho punitivo. Sin embargo, se han presentado diversos proyectos de ley que buscan llenar este vacío normativo; entre ellos destaca la iniciativa para expedir una Ley Federal de Inteligencia Artificial (Senado de la República, 2024) presentada el 02 de abril de 2024 por el senador Ricardo Monreal Ávila, en la cual se proponen establecer principios éticos, mecanismos de supervisión humana, obligaciones de transparencia y criterios de responsabilidad en el uso de sistemas automatizados, particularmente cuando estos afecten derechos fundamentales.

Asimismo, el 10 de diciembre de 2025 el diputado Santiago González Soto, presentó ante la Cámara de Diputados (2025) un proyecto de reforma constitucional que plantea la modificación al artículo 73, con el propósito de otorgarle al Congreso de la Unión la facultad expresa de expedir leyes en materia de IA, para que exista una disposición nacional con términos uniformes, y con ello, evitar que cada estado legisle bajo sus propios criterios. El enfoque del proyecto de reforma se sustenta en la protección de los derechos humanos, la protección de los datos personales, la reducción de brechas de talento y el uso de la IA como herramienta para el bienestar social y económico.

Por otra parte, el 02 de julio de 2025 la extinta Segunda Sala de la Suprema Corte de Justicia de la Nación (2025) emitió un criterio relacionado con el reconocimiento de derechos de autor en una obra realizada con IA, derivado del amparo directo 6/2025. En éste sustenta que cualquier obra realizada con IA, carece de la creatividad y originalidad humana, requisitos *sine qua non* para otorgar una autoría, dado a que, sendos derechos están vinculados con la dignidad y, personalidad de un individuo; atributos que indiscutiblemente no tiene la máquina inteligente. Igualmente, el 09 de enero de 2026, el máximo tribunal mexicano publicó la jurisprudencia denominada *Inteligencia Artificial aplicada en procesos jurisdiccionales. Elementos mínimos que deben observarse para su uso ético y responsable con perspectiva de derechos humanos* (2026), en la cual postulan que los juzgadores que utilicen

herramientas de IA durante el desarrollo de los procesos jurisdiccionales, deben observar como elementos mínimos los principios de: a) proporcionalidad e inocuidad; b) protección de datos personales; c) transparencia y explicabilidad y d) supervisión y decisión humanas.

Conclusiones

La IA se ha posicionado como una de las transformaciones tecnológicas más relevantes para el derecho contemporáneo, al incidir de manera directa en la práctica jurídica, la función jurisdiccional y la formación de los futuros profesionales del derecho. En el desarrollo de la presente colaboración, se ha evidenciado que la IA ofrece importantes beneficios, tales como la optimización de los procesos judiciales, el análisis eficiente de grandes volúmenes de información, la identificación de patrones jurisprudenciales y el apoyo en la toma de decisiones, lo que contribuye a una mayor eficiencia del sistema de justicia; aunque su uso también plantea serios desafíos que no pueden ignorarse, como la utilización de algoritmos predictivos y los modelos automatizados de decisión, que pueden violentar principios fundamentales como el debido proceso, la legalidad, la presunción de inocencia, la motivación de las resoluciones y el derecho a una defensa adecuada. En el mismo sentido, la opacidad de los sistemas algorítmicos, así como la posibilidad de que reproduzcan sesgos históricos derivados de los datos con los que han sido programados, representa un riesgo significativo para las garantías constitucionales y el acceso efectivo a la justicia.

En este sentido, resulta indispensable reconocer que los programas inteligentes no deben concebirse como un sustituto del juez, ni del razonamiento jurídico humano, sino como una herramienta complementaria; puesto que, es claro que, la función jurisdiccional implica interpretación, ponderación de valores, análisis del contexto y sensibilidad frente a las circunstancias particulares de cada caso, elementos y aptitudes que no pueden ser replicados a plenitud por una máquina. Bajo esa premisa, la justicia no puede reducirse a cálculos estadísticos ni a criterios de probabilidad, ya que su esencia radica en la protección de la dignidad humana y los derechos fundamentales.

Actualmente, países europeos como Italia –pionero desde el año 2025- y España –recientemente en el 2026- han apostado en la implementación de una normativa que regule el uso y aplicación de los sistemas inteligentes, tanto en la práctica jurisdiccional, como en áreas diversas; no obstante, sendos parámetros legales aún pueden ser perfectibles. En el caso de Italia, omite adentrarse en el uso de IA en la función jurisdiccional, porque su enfoque está orientado hacia la gobernanza, la soberanía digital y los principios éticos;

a diferencia de España que, se ocupa de regular las hipótesis bajo las cuales aplica o no juzgar con intervención de la IA.

Por su parte, México aún se encuentra en la etapa de definición y presentación de proyectos de reforma que buscan lograr una regulación en materia de IA; aunque la Suprema Corte de Justicia de la Nación ha postulado dos posicionamientos –hasta el momento en el que realiza la presente colaboración- que han sido determinantes para definir el rumbo de la IA en cuestión de derechos de autor, así como en su uso ético y responsable con perspectiva de derechos humanos; todavía queda un largo camino por recorrer y conocer en el tema.

Referencias

Cáceres Nieto, E. (2016). *Tecnicas ericksonianas para la elicitacion del conocimiento judicial en un proyecto de inteligencia artificial aplicada al derecho*. Universidad Nacional Autónoma de Mexico. https://archivos.juridicas.unam.mx/www/bjv/libros/9/4306/1.pdf

Cáceres Nieto, E. (2023). La Inteligencia Artificial aplicada al derecho como una nueva rama de la teoría jurídica. *Anales de la Cátedra Francisco Suárez* (57), 63-89. http://dx.doi.org/10.30827/ACFS.v57i.26281

Cámara de Diputados del H. Congreso de la Unión. (10 de Diciembre de 2025). *Iniciativa que reforma y adiciona el artículo 73 de la Constitución Política de los Estados Unidos Mexicanos, para regular el uso de la Intgeligencia Artificial y otros dispositivos tecnológicos afines*. https://sil.gobernacion.gob.mx/Archivos/Documentos/2025/12/asun_5000350_20251210_1765397459.pdf

Cámara de Diputados y Senado de la República de Italia. (23 de Septiembre de 2025). *Ley número 132. Disposiciones y delegaciones al gobierno en materia de Inteligencia Artificial*. https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:2025-09-23;132

Corte Constitucional República de Colombia. (2020). *Boletín No. 128. La Sala Plena de la Corte Constitucional resaltó la importancia de Pretoria y anunció el inicio de su fase de implementación*. Bogotá. https://www.corteconstitucional.gov.co/noticia.php?PRETORIA,-un-ejemplo-de-incorporaci%C3%B3n-de-tecnolog%C3%A-Das-de-punta-en-el-sector-justicia-8970=

Estevez, E., Linares Lejagarra, S., & Fillotrani, P. (2020). *Prometea: transformando la administración de justicia con herramientas de Inteligencia Artificial*. Buenos Aires: Banco Interamericano de Desarrollo. http://dx.doi.org/10.18235/0002378

Haley, J. (1997). Típicamente Erickson. *Revista de psicoterapia*, *8*(29), 41-62. https://dialnet.unirioja.es/servlet/articulo?codigo=2969890

Inteligencia Artificial aplicada en procesos jurisdiccionales. Elementos mínimos que deben observarse para su uso ético y responsable con perspectiva de derechos humanos, Jurisprudencia 2031640. Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito (12 de Enero de 2026). https://sjf2.scjn.gob.mx/detalle/tesis/2031640

Padilla Sanabria, L. (2025). La inteligencia artificial en el derecho punitivo: riesgos de su aplicación en materia administrativa y disciplinaria. *Ciencia jurídica*, *14*(28), 121-135. https://doi.org/10.15174/cj.v14i28.534

Pleno del Consejo General del Poder Judicial de España. (30 de Enero de 2026). *Instrucción 2/2026, de 28 de enero, sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional*. https://www.poderjudicial.es/cgpj/ca/Temes/Compendi-de-Dret-Judicial-/Instruccions/Instruccion-2-2026—de-28-de-enero—del-Pleno-del-Consejo-General-del-Poder-Judicial—sobre-la-utilizacion-de-sistemas-de-inteligencia-artificial-en-el-ejercicio-de-la-acti

Princi, M. (2024). El espíritu de la ley y el paradigma del poder limitado: el regreso al juez autómata. En A. Abadías Selma, & P. Simón Castellano (Eds.), *Desafíos jurídicos de la Inteligencia Artificial: reflexiones sobre la toma de decisiones judiciales*. J.M. Bosch Editor. https://elibro.net/es/ereader/bibliotecauv/273813?page=1

Quirós Fons, A. (2025). *Propuestas educativas en la era de la IA. Regulación y uso ético*. Dykinson. https://elibro.net/es/ereader/bibliotecauv/285887

Ross Intelligence. (2016). *Meet Ross, your brand new artificially intelligent lawyer. Toronto*. Youtube. https://youtu.be/ZF0J_Q0AK0E?si=ozhW5ssMj44TS139

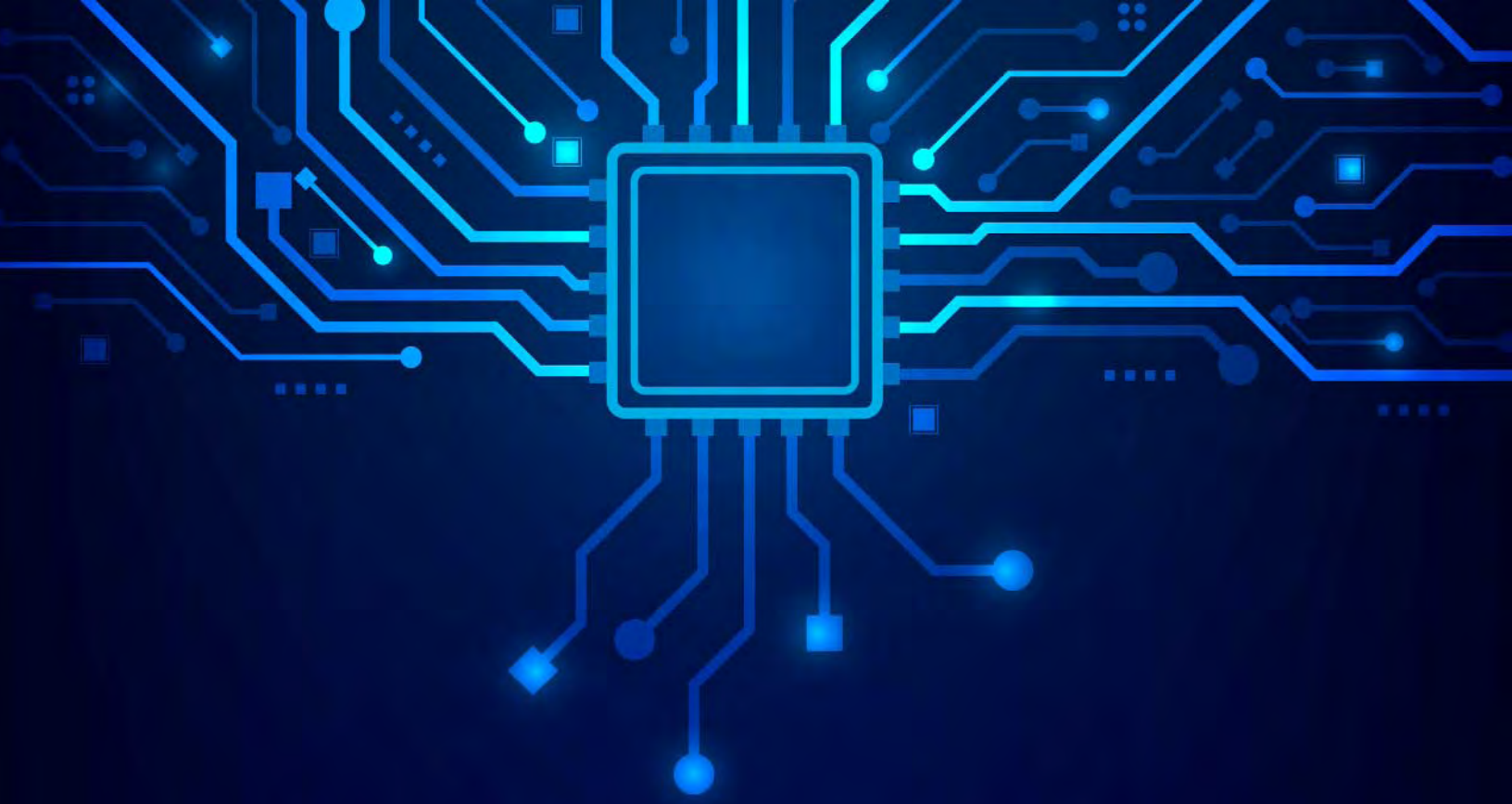
Salas, J. (19 de Enero de 1998). El algoritmo que adivina los delitos futuros falla tanto como un humano. *El País*. https://elpais.com/elpais/2018/01/17/ciencia/1516194073_122982.html

Selma Alfredo Abadias, P. S. (s.f.). *Desafios juridicos de la inteligencia artificial. Reflexiones sobre la toma de deciciones judiaciales*. J. M. Bosch Editor. https://elibro.net/es/ereader/bibliotecauv/273813?page=1

Senado de la República. (02 de Abril de 2024). *Iniciativa con proyecto de decreto por el que se expide la Ley Federal que regula la Inteligencia Artificial*. https://infosen.senado.gob.mx/sgsp/gaceta/65/3/2024-04-02-1/assets/documentos/Inic_Morena_Sen_Monreal_Ley_Fed_Inteligencia_Artificial_02042024.pdf

State of Wisconsin *vs*. Eric L. Loomis, 2015AP000157-CR (State of Wisconsin in Supreme Court 12 de Abril de 2015). https://archive.epic.org/algorithmic-transparency/crim-justice/Loomis-Brief.pdf

Suprema Corte de Justicia de la Nación. (27 de Agosto de 2025). *Sentencia del amparo directo 6/2025*. https://transparencia-ciudadana.scjn.gob.mx/sites/default/files/page/2025-08/scjn-ad-6-2025.pdf



13.

Envejecimiento digno en México: detección en tiempo real del nivel de riesgo de sarcopenia mediante inteligencia artificial

Santiago Arceo-Diaz
Universidad de Colima, México
ORCID: 0000-0002-7085-3653

Xóchitl Trujillo-Trujillo
Universidad de Colima, México
ORCID: 0000-0001-8146-5667

Elena Elsa Bricio-Barrios
Tecnológico Nacional de México / Instituto Tecnológico de Colima, México
ORCID: 0000-0002-1260-9740

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.13

13.

Envejecimiento digno en México: detección en tiempo real del nivel de riesgo de sarcopenia mediante inteligencia artificial

Introducción

Tras la pandemia por COVID-19 quedó claro que contar con espacios adecuados, infraestructura funcional y personal capacitado no es un lujo, sino una prioridad de salud pública (Ramírez et al. 2021). Sin embargo, la eficacia de un sistema de salud no depende únicamente del número de camas o de equipos de última generación, sino de su capacidad para identificar, priorizar y canalizar pacientes de forma oportuna y eficiente (Masfret, 2010).

Los sistemas de cribado o diagnóstico primario, se construyen a partir de la combinación de variables clínicas cuya relación con el diagnóstico no siempre es lineal ni evidente. En enfermedades multifactoriales, los patrones suelen estar distribuidos entre síntomas, antecedentes y mediciones simples, por lo que a simple vista es difícil identificar reglas de decisión consistentes (Ascunce, 2015).

Esta área de oportunidad puede abordarse mediante estrategias de aprendizaje automático, como los árboles de decisión, que permiten jerarquizar variables, identificar puntos de corte y construir reglas interpretables para estimar riesgo en una primera etapa (Martínez et al. 2009). Estos modelos han evolucionado hacia sistemas adaptativos que, además de estimar un diagnóstico primario, permiten que los datos ingresados por el paciente o por el personal de salud, se integren a una base de datos que se actualiza de forma continua. Con ello, el sistema puede reentrenar y ajustar el modelo predictivo con un mayor volumen de información, aumentando su capacidad y precisión de diagnóstico primario (Vega et al. 2020).

Estas herramientas no se limitan a automatizar diagnósticos o reducir la carga de los profesionales, sino a funcionar como un filtro clínico que ayude a priorizar recursos especializados. Este aporte es significativo en México ya que en el año 2022, alrededor del 14%

de la población tenía 60 años o más (casi 18 millones), y se proyecta que para 2050 una de cada cuatro personas será adulta mayor (UNAM, 2023).

Para ofrecer un programa de atención de salud universal, en 2024 se puso en marcha el programa *Salud Casa por Casa* que, a diferencia de los anteriores, propone visitas médicas a domicilio (Secretaria del Bienestar, 2024). Este programa busca que en la primera visita el personal de salud registre signos biométricos y antropométricos para valorar el estado general, complementándolo con cuestionarios validados que identifican autonomía y capacidad cognitiva. Sin embargo, en su estado actual enfrenta un reto de escalabilidad: por cada profesional participante, se estima la visita de nueve hogares diarios, los siete días de la semana.

Entre los problemas de salud que afectan la independencia funcional, la sarcopenia, es una enfermedad crónica ocasionado por una disminución significativa de masa muscular y, en consecuencia baja fuerza y autonomía para realizar actividades cotidianas (Wilkinson et al. 2024). Esta enfermedad, reconocida como un síndrome geríatrico, ha sido diagnosticado clínicamente en personas de 40 a 55 años con historial clínico de hipertensión arterial y diabetes mellitus tipo 2 (Jung et al., 2023).

La Absorciometría Dual de Rayos X (DEXA) y la Impedancia Bioeléctrica (BIA) son los estándares de referencia para calcular el índice de masa masa muscular (IMME), pero su disponibilidad y costo limitan su aplicación como herramienta de detección masiva, especialmente en población económicamente vulnerable. Por ello, cuando se busca atención oportuna a gran escala, incluyendo pacientes jóvenes en riesgo de padecer sarcopenia temprana, se torna indispensable incorporar estrategias de cribado primario que permitan descartar baja probabilidad y canalizar evaluación confirmatoria únicamente en los casos con mayor riesgo (Caicedo et al., 2024).

El diagnóstico primario de sarcopenia ha mostrado alto desempeño con las variables antropométricas y puntos de cohorte establecidas por el Grupo Europeo de Trabajo sobre Sarcopenia en Personas Mayores (EWGSOP), el Grupo Asíatico de Trabajo sobre Sarcopenia en Personas Mayores (AWGS) y el Grupo Internacional de Sarcopenia y Caquexia (SCWD), quienes coincidieron que el IMME, fuerza de prensión palmar, porcentaje de grasa corporal y velocidad de marcha en conjunto son capaces de predecir con alta precisión este síndrome (Sepúlveda et al., 2020). No obstante, estos puntos de corte son exclusivos de su población etaria. Por tanto, es de interés identificar los puntos de cohorte para la población mexicana.

En México, Arceo et al. (2024) reunieron a 150 adultos mayores de centros recreativos de Colima, México, a un estudio aprobado por la Comisión Nacional de Bioética (EICAN-CL08012021-MARCOGP-01). Personal capacitado midió 13 variables antropométricas, incluida BIA, para diseñar un sistema primario de diagnóstico de sarcopenia. Mediante herramientas de inteligencia artificial como *clustering* jerárquico y árboles de decisión identificaron tres sistemas de clasificación basados en variables antropométricas con sus respectivos puntos de corte con una precisión de predicción del 80%.

La siguiente etapa fue compartir esta innovación tecnológica mediante una aplicación web nombrada *SARC-OPEN-IA* (disponible en: https://1introsarcopeniapy-prsmum5dl-bfejjg3mmgspu.streamlit.app/ bajo el Registro Público del Derecho de Autor: 02-2023-101010570900-01). Sin embargo, esta propuesta está limitada por dos factores: a) el tamaño muestral; y, b) los sistemas de clasificación que requieren de IMME, porcentaje de grasa corporal y fuerza de prensión palmar.

A continuación, Arceo et al. (2023) desarrollaron la aplicación web *ANTROPO-FIT* (disponible en: https://happcfsdfvvrtw4zvxmh6z.streamlit.app/ bajo el Registro Público del Derecho de Autor: 03-2024-1213102841100-01). Esta predice el IMME y el porcentaje de grasa corporal a partir del ingreso de una o dos medidas antropométricas, utilizando modelos de regresión lineal, bosque aleatorio y descenso de gradiente. Posteriormente, cada modelo se validó respecto a los diagnósticos por BIA de 62 personas adultas mayores reportando una precisión de diagnóstico del 93%.

Para mejorar la capacidad de predicción del diagnóstico primario de una enfermedad, es necesario contar con más datos para entrenar modelos matemáticos robustos, de manera que esta área de oportunidad se puede abordar mediante modelos de aprendizaje adaptativo (Aparicio & Aparicio, 2024). Estas herramientas presentadas como aplicaciones web y móviles cumplen dos funciones: al ingresar las variables antropométricas, el modelo genera un diagnóstico primario y con los datos ingresados el modelo se actualiza para contar con mayor capacidad de predicción (Hernández & Pérez, 2023).

Por ejemplo, la aplicación *HT Sarcopenia*, creada por el INTA de la Universidad de Chile, estima indicadores de masa muscular mediante medidas corporales como la altura de cadera y el diámetro de pantorrilla, muestra resultados comparables al DEXA y se utiliza en centros de salud. Por su parte, la aplicación *Sit to Stand*, validada en España, analiza en

video el gesto de levantarse de una silla junto con la circunferencia de pantorrilla, logrando alta precisión en la detección de sarcopenia y fragilidad.

Al igual que los modelos predictivos de regresión lineales o factoriales, los modelos supervisados están conformados por métricas de precisión que muestran qué tan bien clasifican a las personas. La exactitud (*accuracy*) resume el porcentaje total de aciertos. La precisión indica qué tan confiables son las predicciones cuando el modelo asigna una categoría, y el *F1-score* resume el equilibrio entre acertar y no dejar casos fuera. El *support* indica cuántos casos hubo en cada categoría. Finalmente, el macro *average* promedia el desempeño dando el mismo peso a cada categoría, mientras que el *weighted average* lo promedia considerando cuántos casos aporta cada categoría (Bobadilla, 2021).

Este trabajo presenta el diseño, evaluación y validación de una aplicación *web* y móvil de libre acceso que ofrece dos funcionalidades: a) consultar variables antropométricas preasignadas o personalizadas junto con los puntos de corte para estimar el riesgo primario de sarcopenia en población mexicana sin comorbilidades; b) ingresar variables o bases de datos para obtener un diagnóstico primario del nivel de riesgo. Además, permite conocer la fiabilidad del modelo adaptativo así como descargar resultados, favoreciendo su accesibilidad y uso en contextos clínicos y de investigación.

Método de investigación

Participantes

Estudio observacional, retrospectivo y analítico basado en registros de derechohabientes del IMSS seleccionados aleatoriamente en las Delegaciones Sur y Norte de la Ciudad de México (2019 y 2022). Los datos fueron proporcionados por el Dr. Sergio Sánchez García, Jefe de la Unidad de Investigación Epidemiológica y de Servicios de Salud, División de Envejecimiento del Instituto Mexicano del Seguro Socia (IMSS).

En este estudio se evaluaron los parámetros antropométricos: talla y peso; circunferencias de pantorrilla, muslo, brazo, cintura y abdomen; y pliegues cutáneos de tríceps, subescapular y pantorrilla (en extremidades se promedió entre lados). Como indicadores de desempeño se consideraron la velocidad de marcha y la fuerza de prensión palmar. De forma complementaria, se contrastaron las mediciones entre 2019 y 2022, periodo coincidente con la emergencia por COVID-19.

El proyecto fue aprobado por el Comité de Bioética del Instituto Estatal de Cancerología y verificado por la Comisión Nacional de Bioética en México, bajo el número de registro CONBIOÉTICA-09-CEI-009-20160601. Participaron 1678 personas en el año 2019 y 751 en 2022, quienes cumplieron con: tener 60 años o más, no presentar amputaciones, no padecer comorbilidades como diabetes mellitus tipo 2, hipertensión arterial o artritis severa, y/o movilidad reducida que requiera el uso de silla de ruedas o condición de postración.

Herramientas

La base de datos en formato .xlsx, se procesó y analizó en *Python* 3.0, un *software* de código abierto. Para ofrecer una interfaz amigable con el usuario y de acceso web, el código se migró a la plataforma gratuita *Streamlit*, que facilita la captura de datos para usuarios sin conocimientos de programación, visualización y descarga de gráficas y variables de salida.

Procedimiento

Dado que esta aplicación web y móvil, nombrada *SARC-PREDICTOR* (disponible en: https://mainpy-lj2lpzdm34zbxhp9swkzba.streamlit.app/ con Registro Público del Derecho de Autor en trámite) fue diseñada con cuatro módulos: a) la presentación de la aplicación; b) el módulo de “Proceso” donde se muestra cómo se identificaron las variables antropométricas y puntos de cohorte; c) el módulo de “Formularios”, donde el usuario ingresa las variables antropométrica o bases de datos y obtiene el diagnóstico primario; y, d) el módulo del equipo de trabajo. A continuación se muestran los detalles del segundo y tercer módulo.

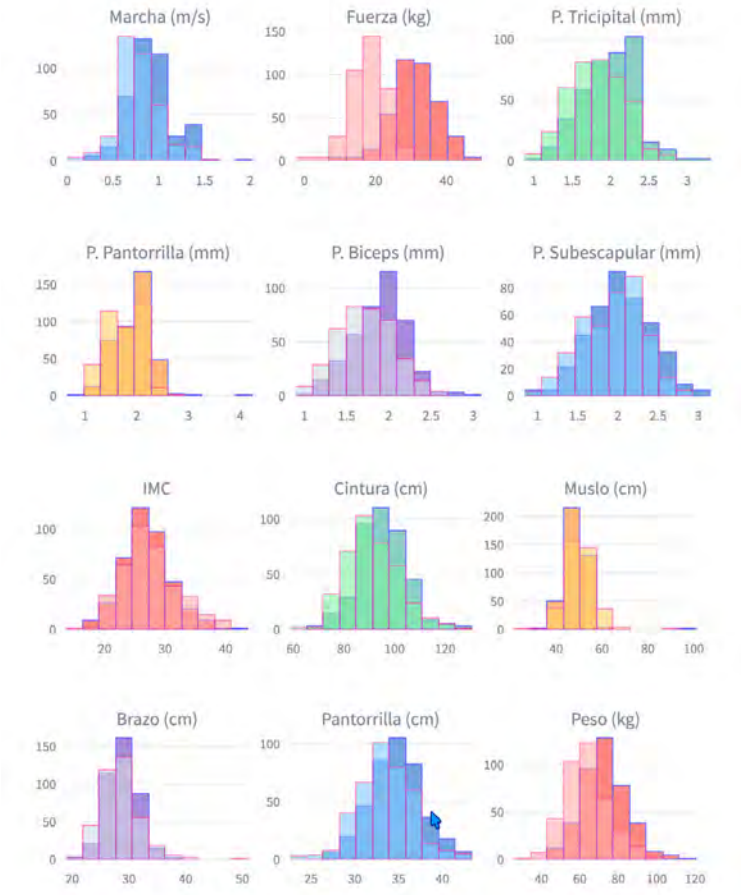
Módulo “Proceso”: La aplicación presenta, la opción de seleccionar el año de referencia (2019 y 2022), correspondientes a los periodos donde se recabaron las respuestas y desplegará los gráficos descriptivos de la muestra (Figura 1), incluyendo la distribución por sexo y un gráfico de barras apiladas, general y por sexo, que muestra la proporción de participantes con comorbilidades asociadas al envejecimiento, tales como diabetes, hipertensión y artritis.

En la Figura 1 se muestra la distribución, por sexo, de las variables anrtropométricas de los pacientes que cumplieron con los criterios de inclusión. Se observa que, en fuerza y circunferencia muscular (brazo, muslo, pantorrilla), los hombres presentan valores más elevados, mientras que las mujeres tienden a concentrarse en rangos más bajos, lo que refleja diferencias importantes en masa y fuerza muscular. En contraste, variables como el índice de masa corporal (IMC), cintura y peso muestran mayor traslape entre ambos sexos,

indicando similitudes en la distribución. En general, hay diferencias marcadas en fuerza y masa muscular y, aunque en menor proporción, en algunas variables corporales. Por tanto, el sexo es un factor determinante en el diagnóstico de IMME.

Figura 1

Histograma de las variables antropométricas más relevantes de los participantes sin comorbilidades del año 2019



La siguiente etapa fue la generación de los modelos predictivos, el primer paso consistió en analizar aquellas personas sin comorbilidades, con el fin de aislar la relación entre las variables antropométricas y las predicciones y evitar que los efectos de dichas condiciones sesgaran los resultados.

Para disponer de un modelo robusto, sin perder la precisión, se identificaron los parámetros antropométricos más relevantes mediante el cálculo de la varianza normalizada, seleccionando aquellas variables con una magnitud superior a 0.01. Posteriormente, se aplicó

la prueba de correlación de Pearson para determinar qué parámetros estaban fuertemente asociados entre sí, de manera que, en lugar de realizar dos o más mediciones redundantes, fuese suficiente con registrar solo una.

El cuadernillo no incluye mediciones directas del IMME de los pacientes, un indicador central para el diagnóstico de sarcopenia, por lo que se desarrolló un módulo que lo calcula a partir de múltiples combinaciones de parámetros antropométricos. Las estimaciones se compararon con la fórmula validada de Ramírez et al. (2015), encontrando que algunos de nuestros modelos tenían un error medio de 0.06 respecto a sus predicciones pero que ofrecían la ventaja de poder calcular una predicción a partir de las variables disponibles para cada caso.

A continuación, se identificaron los niveles de riesgo de sarcopenia de la población mexicana. Para ello, se retomaron las variables propuestas por EWGSOP, AWGS y SCWD, conformadas por la fuerza de prensión palmar, el IMME y la velocidad de marcha (Sepúlveda et al., 2020). Con estas variables se aplicó un algoritmo de clúster jerárquico secuencial de tres niveles.

Siguiendo la metodología reportada por Arceo et al. (2026), se realizó una categorización de los niveles de riesgo con base en el segundo quintil de cada participante. Por ejemplo, las personas saludables presentaron fuerza de prensión por encima del segundo quintil. Con este criterio, se definieron cuatro categorías: saludables (fuerza ≥ 2.º quintil), sarcopenia sospechosa (fuerza < 2.º quintil), sarcopenia probable (fuerza e IMME < 2.º quintil) y sarcopenia grave(fuerza, IMME y velocidad de marcha < 2.º quintil).

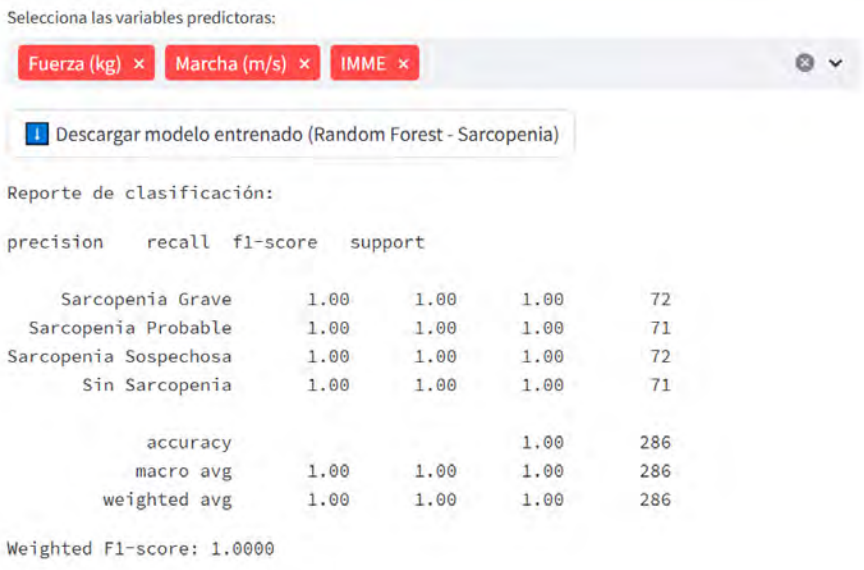
Como ejemplo representativo del módulo de Proceso, se seleccionaron las variables antropométricas recomendadas por el EWGSOP, fuerza de prensión palmar, velocidad de marcha e IMME, por su valor para discriminar el nivel de riesgo de sarcopenia. La Figura 2 presenta las métricas de desempeño del modelo adaptativo en la muestra que cumplió los criterios de inclusión.

La Figura 2 muestra el reporte de clasificación del modelo para cuatro categorías: Sarcopenia grave, sarcopenia probable, sarcopenia sospechosa y sin sarcopenia. En las cuatro, el modelo obtuvo valores perfectos en los indicadores de desempeño, lo que significa que clasificó correctamente todos los casos en la muestra evaluada. Además, se observa que el número de participantes por categoría fue muy similar (71 a 72 casos por grupo), sumando

un total de 286 registros. En conjunto, la exactitud global (accuracy) también fue 1.00, y los promedios macro y ponderado (weighted) reportan el mismo resultado, consistente con un desempeño uniforme en todas las categorías.

Figura 2

Captura de pantalla de un reporte sobre la precisión del modelo para riesgo de sarcopenia

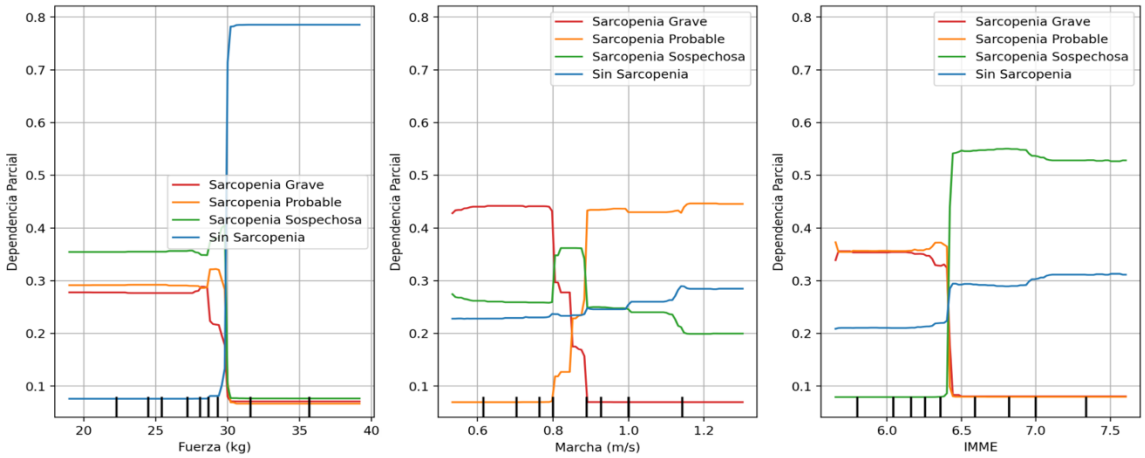


Al comparar el desempeño de esta propuesta respecto a modelos de regresión lineal empleando el perímetro de cintura, Guillot et al. (2024) reportaron una baja correlación (r = 0.31), mientras que Shi et al. (2022) lograron una correlación mucho mayor (r = 0.91) aplicando técnicas avanzadas de *machine learning* con múltiples variables antropométricas. En contraste, este trabajo permite al usuario seleccionar variables antropométricas que tenga disponibles y dependiendo de las variables de entrada mostrará el IMME y riesgo de sarcopenia con su respectivo error.

Finalmente, estos modelos predictivos se construyen a partir de los puntos de corte, los cuales se obtuvieron mediante las gráficas de dependencia parcial, que muestran cómo cambia la predicción del modelo cuando una variable aumenta o disminuye, manteniendo las demás constantes. De este modo, es posible identificar rangos donde la probabilidad de riesgo se incrementa de forma marcada y definir umbrales prácticos para clasificar a los participantes. La Figura 3 muestra los puntos de corte con los cuatro niveles de riesgo de padecer sarcopenia de las variables empleadas del ejercicio previo.

Figura 3

Gráficas de dependencia parcial para la clasificación del nivel de riesgo de sarcopenia de acuerdo a las variables antropométricas elegidas por el usuario



La Figura 3 muestra los umbrales de los niveles de riesgo de sarcopenia. En general, se observa que los umbrales son consistentes para la clasificación. “Sin sarcopenia” se asocia principalmente con fuerza por encima de alrededor de 30 kg, con contribución adicional de mayor velocidad de marcha. “Sospechosa” se relaciona con fuerza baja, pero con IMME relativamente conservado (incremento marcado en 6.5). Mientras que el nivel “Probable” aumenta con fuerza baja y valores de IMME por debajo de ~6.5, con un aporte relevante de marcha en el rango de 0.85 a 0.9 m/s. Finalmente, “grave” se concentra cuando convergen fuerza baja, IMME bajo y marcha lenta, con un descenso abrupto al superar entre 0.8 a 0.85 m/s.

Una vez establecida la fiabilidad del modelo adaptativo, el usuario puede ingresar las variables antropométricas disponibles para obtener como salida el nivel de riesgo primario de sarcopenia. Esto se realiza en el módulo “Formularios”, donde se estima el IMME a partir de la selección de 2, 3 o 4 variables con mayor capacidad predictiva (según los registros 2019 y 2022). Si el usuario no cuenta con dichas variables, puede ingresar las que tenga disponibles y el sistema reporta la estimación de IMME junto con su error de predicción. Finalmente, el módulo permite generar y descargar el diagnóstico primario del nivel de riesgo.

A modo de ejemplo, en la Tabla 1 se presenta el caso de una mujer adulta mayor mexicana con estatura de 150 cm, peso de 61 kg e IMC de 27 kg/m². Además, se incluyen la circunferencia de pantorrilla de 30 cm y fuerza de prensión palmar de 17 kg, cuyos puntos de corte están en el umbral de padecer sarcopenia (Sepúlveda et al., 2020), obteniendo diferentes niveles de riesgo.

Tabla 1

Relación de variables antropométricas, IMME estimado, error del modelo y diagnóstico primario del nivel de riesgo de sarcopenia

Cantidad de variables	Variable antropométricas	IMME estimado	MSE	Nivel de riesgo de sarcopenia
1	Sexo	5.79	0.3524	Probable
2	Sexo, IMC	5.52	0.1885	Probable
3	Sexo, IMC, Peso	5.51	0.1166	Grave
4	Sexo, IMC, Estatuta, Pantorrilla	5.21	0.0894	Grave
5	Sexo, IMC, Estatura, Pantorrilla y Fuerza de prensión palmar	5.21	0.0649	Grave

La Tabla 1 muestra que, conforme se incrementa la cantidad de variables antropométricas, el modelo mejora su capacidad de estimación de IMME, reflejado en una reducción progresiva del MSE (de 0.3524 con 1 variable a 0.0894 con 4 variables). Sin embargo, se observa un punto de estabilización: a partir de cuatro variables, el IMME estimado permanece sin cambios (5.21), y la incorporación de una quinta variable no reduce el error de forma significativa.

Además, esta estabilidad es consistente con el diagnóstico primario de riesgo de sarcopenia: con 1 a 2 variables el modelo clasifica como sarcopenia probable, mientras que con 3 o más variables la clasificación se consolida como sarcopenia grave y se mantiene aún al añadir más información. En comparación con el reporte EWGSOP para mujeres, los valores de IMME estimados muestran un patrón relevante: con 2 a 3 variables el IMME se ubica cerca del umbral típico de baja masa muscular, mientras que con 4 a 5 variables desciende a 5.21, muestra la importancia de la medición de esta variable como principal predictor de sarcopenia.

Conclusiones

Este trabajo reporta cómo la salud pública puede beneficiarse del uso de la ciencia de datos. A partir de la depuración y análisis de las respuestas recolectadas en el *Cuadernillo de Obesidad, Sarcopenia y Fragilidad en Adultos Mayores del IMSS* en el año 2019 y 2022, se diseñó y desarrolló un sistema adaptativo de diagnóstico primario de niveles de riesgo de sarcopenia en la población mexicana.

Para acercar esta innovación tecnológica al ámbito de la salud pública, se desarrolló una aplicación *web* móvil en tiempo real que permite al usuario consultar las variables antropométricas más relevantes o elegir las que tenga disponible, ya sea de forma manual o cargando un archivo .csv para obtener un diagnóstico primario.

Esta aplicación *web* móvil ofrece una alternativa que solo requiere un flexómetro y una báscula, con un costo aproximado de 350 pesos mexicanos, además, y puede integrarse en programas de atención comunitaria para la detección temprana de sarcopenia para favorecer un envejecimiento digno en México.

Referencias

Aparicio-Gómez, O. Y., & Aparicio-Gómez, W. O. (2024). Innovación educativa con sistemas de aprendizaje adaptativo impulsados por Inteligencia Artificial. *Revista Internacional de Pedagogía e Innovación Educativa*, *4*(2), 343–363.

Arceo-Díaz, S., Bricio-Barrios, E. E., Trujillo-Trujillo, X. A. R., González-Farías, J. R., Bricio-Barrios, J. A., Rios-Silva, M., & Huerta-Viera, M. (2023). Primary screening system for sarcopenia in elderly people based on artificial intelligence. *Revista Mexicana de Ingeniería Biomédica*, *44*(spe1), 53–69.

Arceo-Díaz, S., Bricio-Barrios, E. E., González-Farías, J. R., De la Madrid-Trevizo, L. S., Trujillo-Trujillo, X. A., Bricio-Barrios, J. A., … y Huerta-Viera, M. (2024). Sistema biomédico basado en inteligencia artificial para estimar indirectamente sarcopenia en personas adultas mayores mexicanas. *Revista ELECTRO*, *46*(1), 18–24.

Arceo-Díaz, S., Bricio-Barrios, E. E., Trujillo-Trujillo, X. A., Sánchez-García, S., Bricio-Barrios, J. A., Rios-Silva, M., & Huerta-Viera, M. (2026).Predicted models of anthropometric parameters for primary screening of sarcopenia based on Machine Learning. *Revista Mexicana de Ingeniería Biomédica*, *47*(1).

Ascunce Elizaga, N. (2015). Cribado: para qué y cómo. *Anales del Sistema Sanitario de Navarra*, *38*(1), 5–7.

Bobadilla, J. (2021). *Machine learning y deep learning: usando Python, Scikit y Keras*. Ediciones de la U.

Caicedo Correa, S. M., Aruachán, S. A., Chavarro Carvajal, D. A., Heredia Ramírez, R. A., Moreno, M. M., Llano, C. C., … y Chimbí, C. M. (2024). *Abordaje geriátrico integral: Escalas de uso frecuente en geriatría*. Pontificia Universidad Javeriana.

Gaceta UNAM. (2023). *México, oficialmente un país envejecido*. https://www.gaceta.unam.mx/mexico-oficialmente-un-pais-envejecido/

Guillot, C. C., Neri, L. E., & Aguilar, I. S. (2024). Predicción de probabilidad de índice de masa corporal a partir de la circunferencia de cintura. *Revista Cubana de Investigaciones Biomédicas*, *43*(1), 1–10.

Hernández, M. R., & Pérez, F. L. (2023). M-learning como herramienta para el aprendizaje adaptativo. *Revista Internacional de Humanidades*, *12*(3), 1–15.

Jung, H. N., Jung, C. H., & Hwang, Y. C. (2023). Sarcopenia in youth. *Metabolism*, 144, 155557.

Martínez, R. E. B., Ramírez, N. C., Mesa, H. G. A., Suárez, I. R., Trejo, M. D. C. G., León, P. P., & Morales, S. L. B. (2009). Árboles de decisión como herramienta en el diagnóstico médico. *Revista Médica de la Universidad Veracruzana*, *9*(2), 19–24.

Masfret, D. C. (2010). El trabajo social sanitario en el marco de la optimización y sostenibilidad del sistema sanitario. *Revista de Servicios Sociales*, (47), 109–119.

Ramírez, F. B., Misol, R. C., Alonso, M. D. C. F., & Tizón, J. L. (2021). Pandemia de la COVID-19 y salud mental: reflexiones iniciales desde la atención primaria de salud española. *Atención Primaria*, *53*(1), 89–101.

Ramírez, M. C., Enríquez-Reyna, G., Garza-Sepúlveda, A., Tijerina-Sáenz, E., Ramos-Peña, E., & Gómez de la Garza, M. (2015). Puntos de corte y validación de una ecuación antropométrica para estimar la masa muscular, en el estudio de la sarcopenia en población mexicana. *Salud Pública de México*, *57*(6), 485–486.

Secretaría de Bienestar, Gobierno de México. (2024). *Programa Salud Casa por Casa*. https://programasparaelbienestar.gob.mx/salud-casa-por-casa/

Sepúlveda-Loyola, W. A., Luna Corrales, G. A., Ganz, F., González Caro, H., &y Probst, V. S. (2020). Sarcopenia, definición y diagnóstico: ¿Necesitamos valores de referencia para adultos mayores de Latinoamérica? *Revista Chilena de Terapia Ocupacional*, *20*(2), 259–267.

Shi, J., He, Q., Pan, Y., Zhang, X., Li, M., & Chen, S. (2022). Estimation of appendicular skeletal muscle mass for women aged 60–70 years using a machine learning approach. *Journal of the American Medical Directors Associatio*n, *23*(12), 1985. e1–1985.e8.

Vega, M. Á., Mora, L. M. Q., & Badilla, M. V. C. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, *5*(8), e557–e557.

Wilkinson, I., Raine, T., Wiles, K., Hateley, P., Kelly, D., & McGurgan, I. (2024). *Oxford Handbook of Clinical Medicine*. Oxford University Press.

14.

El impacto de la inteligencia artificial generativa en la arquitectura editorial y el crecimiento exponencial de la producción científica en salud

Salvador Ortiz Santos
Universidad Autónoma de Querétaro, México.
ORCID: 0000-0001-9383-0705

Georgina del Carmen Mota Valtierra
Universidad Autónoma de Querétaro, México.
ORCID: 0000-0001-5856-8633

Humberto Aguirre Becerra
Universidad Autónoma de Querétaro, México.
ORCID: 0000-0003-4093-7810

Blanca Cecilia López Ramírez
Tecnológico Nacional de México /IT de Roque, México
ORCID: 0000-0001-6403-6641

Ma. Cristina Vázquez Hernández
Tecnológico Nacional de México /IT de Roque, México
ORCID: 0000-0002-2647-5464

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.14

14.

El impacto de la inteligencia artificial generativa en la arquitectura editorial y el crecimiento exponencial de la producción científica en salud

Introducción

La producción de conocimiento científico en el campo de la salud experimenta una transformación estructural sin precedentes en la tercera década del siglo XXI (De Vere Hunt et al., 2025). Tradicionalmente, la publicación de investigación dependía de procesos de validación lentos, anclados en la revisión por pares humana y en la selección exhaustiva de contenidos por parte de editoriales de élite que actuaban como los únicos custodios de la verdad científica (Bouderhem, 2024). No obstante, la articulación entre la digitalización a gran escala y la inteligencia artificial (IA) introduce una ruptura sustantiva con el modelo histórico previo (Shonubi, 2026). El problema ya no se limita al aumento de la producción científica, sino a la acumulación acelerada de información y a la insuficiencia de los enfoques tradicionales para manejar volúmenes de datos que crecen de manera exponencial y desbordan la capacidad del juicio humano para procesarlos de forma efectiva (Raman et al., 2024). En este escenario, la presente investigación cobra sentido al examinar el papel de la IA como un vector transversal de transformación, capaz de dinamizar los distintos ámbitos de la salud, desde el trabajo experimental en laboratorio hasta la circulación pública del conocimiento.

La manera en que se comunica la ciencia está cambiando de forma visible. Lo que durante años creció de manera relativamente ordenada hoy se ha convertido en una expansión rápida y más difícil de controlar, primero impulsada por la crisis sanitaria del 2020 y después por la llegada de la inteligencia artificial generativa (IAG), a finales de 2022. Este cambio no solo ha marcado el ritmo al que se produce el conocimiento, sino que también ha dejado huellas duraderas en la productividad científica a escala global (Maupin et al., 2025). Continuando bajo el mismo contexto, las megarrevistas no solo han ganado visibilidad, sino que se han vuelto piezas clave del actual sistema de publicación científica.

Con ello, el peso que antes tenía la exclusividad editorial ha ido perdiendo terreno frente a una dinámica dominada por la urgencia de publicar, el acceso abierto y la estandarización de los procesos técnicos (Liu et al., 2023). A la par, las plataformas digitales dependen cada vez más de herramientas algorítmicas para gestionar volúmenes de manuscritos que hace apenas unos años habrían sido inmanejables, privilegiando la rapidez en la difusión de resultados sobre los tiempos, tradicionalmente más largos, del arbitraje clásico (Maupin et al., 2025). Todo esto deja una pregunta de fondo que no es menor: ¿de qué manera estas tecnologías están redistribuyendo la influencia y reconfigurando el liderazgo científico a escala global?

La arquitectura editorial tradicional, consolidada durante el siglo XX, operaba bajo un modelo de *escasez gestionada*. Las revistas de élite limitaban el espacio físico y la disponibilidad de revisores, lo que garantizaba un filtro estricto, pero a menudo excluyente. La irrupción de la IAG ha convertido esa escasez en una abundancia inmanejable. El crecimiento de manuscritos en salud, facilitado por herramientas que asisten en la redacción de informes clínicos y artículos de revisión, ha llevado al límite el sistema de *peer review*. Según Maupin et al. (2025), el tiempo promedio que un experto dedica a revisar un artículo ha disminuido un 30% en los últimos años debido a la presión del volumen. En respuesta, las editoriales han implementado una *arquitectura asistida* donde la IA realiza controles de plagio, detecta inconsistencias en datos experimentales y sugiere revisores automáticos. Si bien esto permite procesar miles de artículos mensualmente, autores como Liu et al. (2023) advierten que este desplazamiento hacia el algoritmo genera una *caja negra* editorial donde los criterios de aceptación pueden volverse opacos.

A nivel geopolítico, la investigación evidencia una descentralización del poder científico hacia economías emergentes con gran capacidad tecnológica (Reddy, 2024). Las potencias tradicionales de Occidente pierden terreno frente a naciones que invierten estratégicamente en infraestructura digital y formación de talento humano. Este giro histórico demuestra que la ciencia contemporánea es más diversa, multipolar y tecnodependiente que a inicios del siglo XXI (Adarkwah et al., 2025). China, por ejemplo, ha consolidado su posición como líder mundial en cuanto al volumen de producción científica en medicina asistida por IA. Según Xie et al. (2025), la inversión estratégica en clústeres de supercomputación y bases de datos genómicas masivas ha permitido que los investigadores chinos publiquen a un ritmo 2.4 veces superior al de sus pares estadounidenses en áreas de diagnóstico por imagen y esta ventaja no es solo cuantitativa; la arquitectura editorial china ha integrado la IA generativa

desde las etapas de experimentación, reduciendo el ciclo de *laboratorio a publicación* en un 40% (Adarkwah et al., 2025).

Resulta fundamental analizar la integración de países como México e India en las redes globales de innovación. India ha emergido como el centro neurálgico de la democratización de la IA en salud, empleando modelos de lenguaje de gran escala (LLMs) ajustados a contextos locales para liderar la producción en salud pública y telemedicina (Reddy, 2024). Por su parte, México ha mostrado un crecimiento sostenido en la integración de redes de innovación digital y, aunque enfrenta el desafío de la *fuga de datos* hacia infraestructuras tecnológicas extranjeras (Adarkwah et al., 2025), la dependencia de servidores y modelos propietarios de *Big Tech* plantea interrogantes necesarias sobre la equidad tecnológica global y la soberanía científica de las naciones en desarrollo.

La integración de la IA como un motor transversal en las disciplinas de la salud que conlleva implicaciones éticas y estructurales que requieren vigilancia constante (Bouderhem, 2024). El uso de modelos generativos optimiza procesos críticos como la síntesis de imágenes, el diagnóstico predictivo y la redacción de informes técnicos complejos (Zaghir et al., 2024), una de las tendencias más disruptivas es el uso de IA para generar *pacientes sintéticos*. En la investigación farmacológica, los modelos generativos crean conjuntos de datos que imitan características biológicas reales sin comprometer la privacidad, lo anterior, acelera la producción científica al reducir la necesidad de reclutamientos físicos lentos (Zaghir et al., 2024). No obstante, Shonubi (2026) sugiere que entramos en la era de la *Ciencia Sintética*, donde la arquitectura editorial deberá incluir protocolos de verificación de identidad de datos basados en *blockchain* para distinguir hallazgos empíricos de simulaciones.

La síntesis de evidencia médica, pilar de la práctica clínica, también ha sido revolucionada; lo que antes tomaba meses de revisión sistemática hoy se lleva a cabo en horas mediante agentes de IA que analizan metadatos de millones de artículos (Raman et al., 2024), esto ha provocado un crecimiento exponencial en la publicación de revisiones bibliográficas y metaanálisis, cuyo resultado ha saturado las bases de datos como *PubMed,* con información que, a menudo, carece de la profundidad del juicio clínico humano (Hu et al., 2025). Emerge aquí el riesgo de una dependencia tecnológica que podría erosionar el juicio crítico del profesional de la salud. Y así, en consecuencia, la literatura actual enfatiza la necesidad de un marco de competencias transversales que asegure el uso responsable y ético de estas herramientas (Bouderhem, 2024; Mirzaei et al., 2024; Hu et al., 2025).

El riesgo de las *alucinaciones algorítmicas* (generación de datos falsos, pero plausibles) es particularmente grave en salud. Hu et al. (2025) documentan casos donde referencias bibliográficas inexistentes superaron filtros editoriales automatizados y, además, los sesgos inherentes a los datos de entrenamiento de la IA, que a menudo subrepresentan a poblaciones del Sur Global, que pueden perpetuar inequidades en los protocolos de tratamiento sugeridos en la literatura científica (Mirzaei et al., 2024). La ética de la publicación ya no se limita a evitar el plagio, sino a gestionar la agencia intelectual, donde la arquitectura editorial del futuro exigirá que los autores declaren explícitamente el porcentaje de contribución algorítmica, pasando de la *IA como herramienta* a la *IA como colaborador responsable* bajo supervisión humana (Bouderhem, 2024).

La consolidación de la IA generativa ha alterado la arquitectura editorial en salud de forma irreversible e impulsando una producción científica que crece a un ritmo exponencial. Si bien esto democratiza el acceso, también amenaza con colapsar los sistemas de validación. Por otro lado, el liderazgo geopolítico se desplaza hacia quienes dominan la infraestructura de la IA, pero la verdadera frontera será la capacidad de mantener el juicio humano y la integridad en un ecosistema saturado de datos sintéticos. De esta manera, el éxito de esta transformación dependerá de nuestra capacidad para rediseñar las instituciones editoriales, asegurando que la tecnología sea un motor de progreso real y no solo una maquinaria de producción masiva de papel digital.

Para el análisis multidimensional la investigación de este trabajo se fundamenta en las siguientes interrogantes clave: ¿Cómo ha evolucionado el volumen de producción médica y qué hitos marcaron su inflexión desde el año 2000? ¿De qué manera la IA generativa impacta de manera transversal en las disciplinas relacionadas con la salud? ¿Cómo ha cambiado el modelo de difusión del conocimiento a través de las megarrevistas de acceso abierto? ¿Cuál es la configuración actual del mapa geopolítico científico y cómo evolucionan las economías emergentes? Finalmente, ¿qué implicaciones éticas supone la consolidación de la inteligencia artificial como el motor en investigación contemporánea?

Método de investigación

La información se obtuvo de la base de datos *Scopus* con una investigación cuantitativa de carácter bibliométrico, donde la muestra se delimitó y se diseñó una estrategia de búsqueda basada en descriptores clave dentro de los campos de título, resumen y palabras clave: como TITLE-ABS-KEY (“medicine” OR “biomedical” OR “clinical research” OR “healthcare”

OR “health care” OR “drug discovery” OR “medical imaging” OR “diagnostics”). El alcance temporal de la recolección de datos se delimitó entre los años 2000 y 2025, fueron descargados tres conjuntos de datos específicos desde las herramientas de análisis de *Scopus*: de producción por país (*Scopus-Analyze-Country*), de métricas por fuente o revista (*Scopus-Analyze-Source*) y también de distribución por áreas temáticas (*Scopus-Analyze-Subject*).

Procesamiento y análisis de datos en RStudio

En el procesamiento de los registros fueron consolidados en archivos de hoja de cálculo y se procesaron mediante el software estadístico *RStudio*, debido a la estructura no lineal de los reportes originales donde los datos se organizaron en bloques horizontales por año, se implementó un flujo de trabajo computacional basado en los siguientes pasos técnicos:

1. Reestructuración dinámica: se desarrollaron funciones iterativas para procesar la matriz de datos por pares de columnas (bloques), en este procedimiento permitió extraer el año y la métrica correspondiente, transformando los datos de un formato ancho (*wide*) a uno largo (*long*) apto para análisis estadístico.

2. Normalización y limpieza: las técnicas de limpieza de texto se aplicaron mediante expresiones regulares (*gsub*) para estandarizar los años y la función *trimws* para eliminar espacios en blanco en los nombres de países, fuentes y áreas, garantizando la integridad de los filtros comparativos.

3. Análisis de series temporales: se gestionó la variabilidad de los datos para fuentes con aparición tardía (como *Scientific Reports* o *Plos One*), permitiendo que las líneas de tendencia reflejaran el año exacto de inicio de indexación sin distorsionar la continuidad de las fuentes históricas.

4. Visualización, las tendencias de crecimiento y los cambios en el liderazgo editorial y geográfico se representaron mediante gráficos de líneas multivariantes utilizando la librería *ggplot2*, empleando escalas temporales continuas y rotación de etiquetas para optimizar la legibilidad de la cronología.

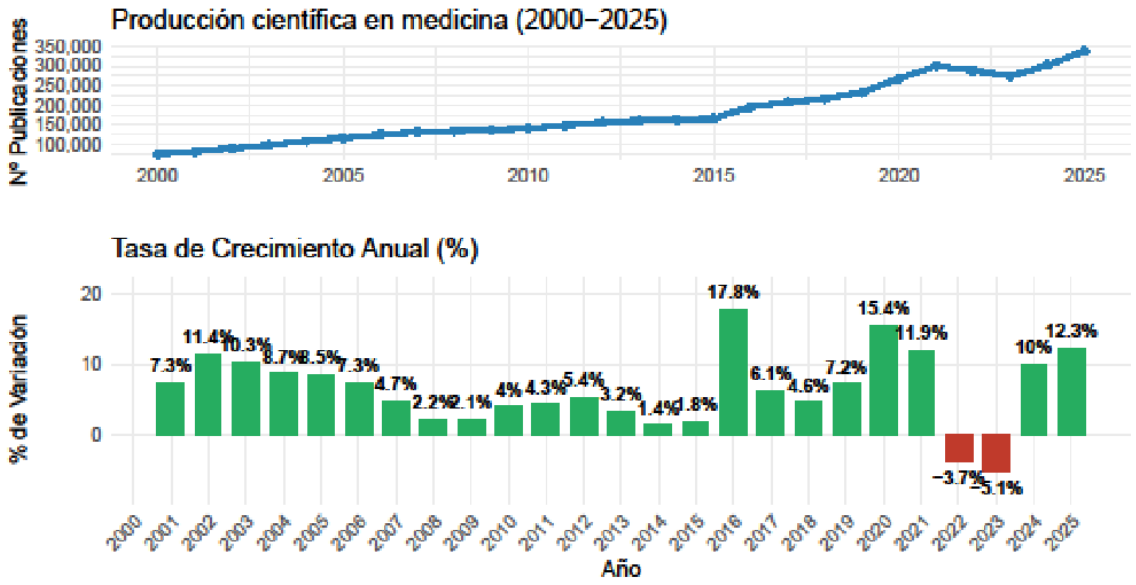
Resultados

Como se observa en la Figura 1, la producción científica en medicina, desde la clínica hasta la farmacéutica, ha experimentado transformaciones significativas. Este avance inició con una etapa de estabilidad entre 2000 y 2006, donde la digitalización y el acceso a internet

sostuvieron crecimientos superiores al 7%. Un hito clave fue 2003, cuando la culminación del proyecto del genoma humano impulsó la medicina molecular con un repunte del 11.4%. Tras una desaceleración con un mínimo de 1.4% en 2014, el sector vivió un cambio de paradigma en 2016 con un salto del 17.84%, motivado por la integración de la IA y el *Big Data*. Posteriormente, la crisis sanitaria de 2020 elevó la producción un 15.37%, alcanzando un pico de 299,823 publicaciones en 2021. No obstante, el bienio 2022-2023 marcó una anomalía con tasas negativas (-3.7% y -5.1%) debido al fin de la urgencia pandémica y a criterios editoriales más estrictos. Finalmente, entre 2024 y 2025 se registró una recuperación robusta alcanzando el 12.35%, cerrando con un récord de 338,690 publicaciones. Este repunte consolida a la inteligencia artificial generativa como el motor principal de una nueva era de crecimiento en la medicina.

Figura 1

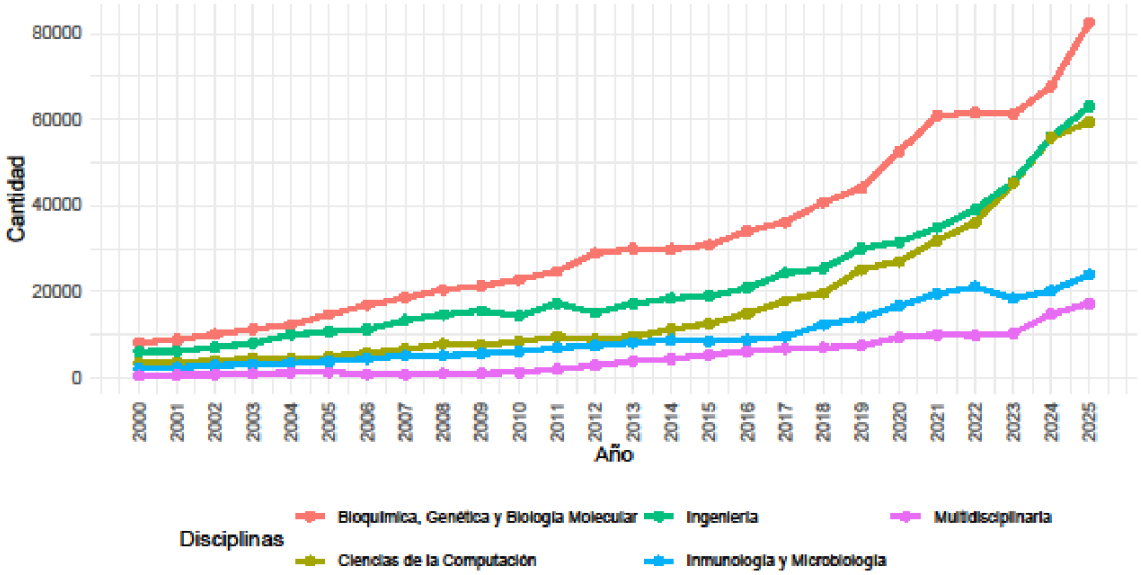
Publicaciones en la disciplina de medicina (Figura propia con datos de SCOPUS)



La Figura 2 presenta un análisis comparativo entre diversas áreas, incluyendo ingeniería, ciencias de la computación, inmunología y bioquímica, entre otras. Si bien la producción científica ha mantenido una tendencia ascendente sostenida desde el año 2000, se aprecia una marcada aceleración en todas las disciplinas durante el bienio 2024-2025. Este repunte exponencial coincide con el auge de la IA generativa, que ha impulsado la investigación en múltiples campos del conocimiento.

Figura 2

Comparativa de disciplinas en investigación de la salud (Figura propia con datos de SCUPOS)



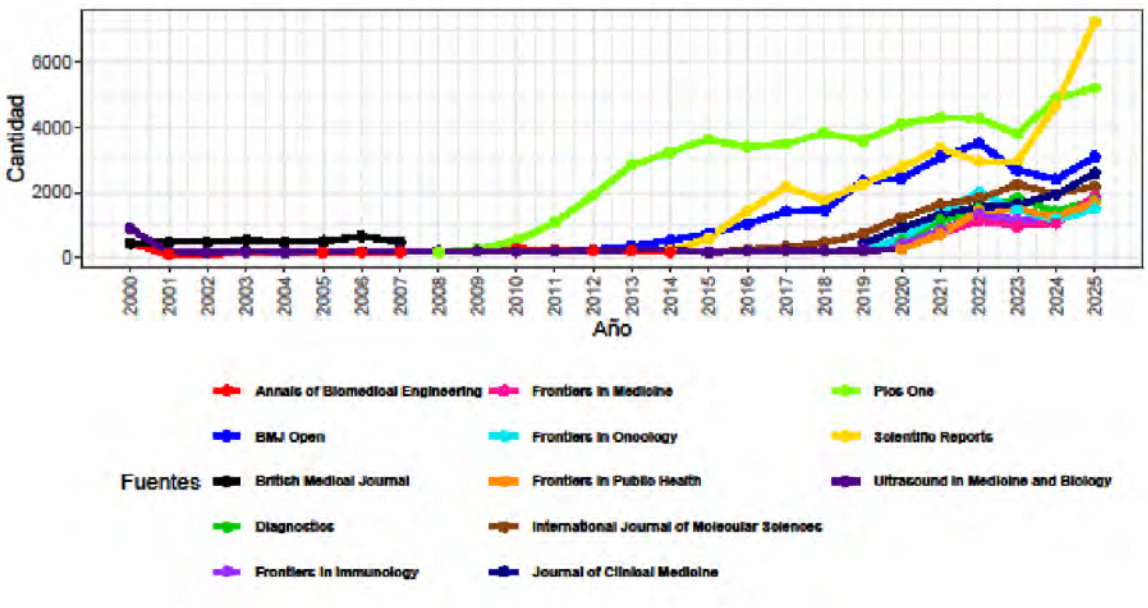
El análisis de la producción científica por revista revela un cambio de paradigma en la difusión del conocimiento médico y biotecnológico. En el año 2000, el liderazgo pertenecía a revistas tradicionales de alto prestigio, como *British Medical Journal* (BMJ) y *Ultrasound in Medicine and Biology*; éstas mantenían volúmenes estables de entre 400 y 900 publicaciones anuales dentro del área de interés (Figura 3). A partir del 2008, la irrupción de *Plos One* transformó el modelo editorial al introducirla evaluación por rigor técnico, el acceso abierto total y el concepto *megarrevista*, escalando de 181 publicaciones en su debut a más de 3,000 en 2014, consolidándose como el principal destino de la producción científica en los temas de interés. La rentabilidad y eficiencia de este modelo impulsó a las editoriales tradicionales a crear sus propias megarrevistas para competir, resultando en el lanzamiento de *Scientific Reports* (2012) y *BMJ Open* (2015).

Entre 2020 y 2021, impulsadas por la crisis sanitaria, estas revistas alcanzaron picos históricos; *Plos One*, por ejemplo, registró 4,318 publicaciones. Asimismo, este periodo consolidó la relevancia de nuevos modelos y revistas que ganaron terreno desde 2015, como el *International Journal of Molecular Sciences* y el ecosistema *Frontiers*, cuya presencia destacó a partir de 2019.

Tras un ligero ajuste en 2023 derivado del fin de la pandemia (periodo en el que se aprecia una reducción en los volúmenes de casi todas las revistas). El bienio 2024-2025 muestra un repunte sin precedentes, donde destaca *Scientific Reports*, que en 2025 alcanzó las 7,246 publicaciones, superando a *Plos One* en el campo de interés. Este crecimiento final confirma la consolidación de plataformas multidisciplinarias que, mediante procesos digitales y la IAG, gestionan volúmenes de información inalcanzables para el modelo editorial del año 2000.

Figura 3

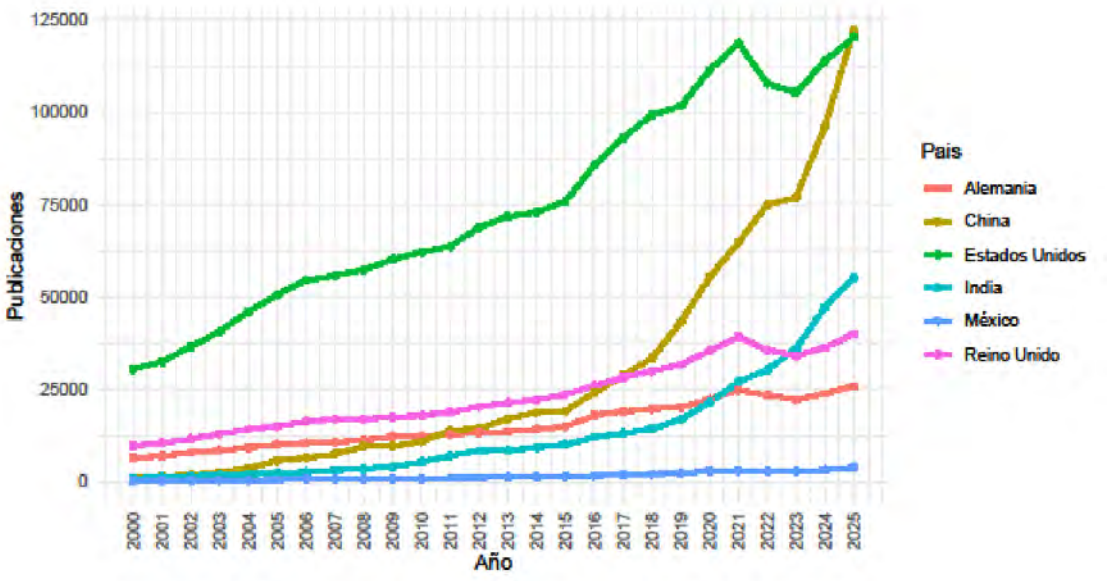
Publicaciones por revistas en temas de la salud (Figura propio con datos de SCOPUS)



Desde el año 2000 (Figura 4), la producción científica mundial ha transitado por tres fases: expansión, crisis y consolidación digital. En la etapa inicial, Estados Unidos y Europa lideraban con amplitud; en contraste, México apenas superaba las 300 publicaciones anuales. Sin embargo, la irrupción de las plataformas digitales facilitó un crecimiento exponencial que culminó en 2025 con China a la cabeza del *ranking* mundial en temas de salud. En este trayecto, México ha experimentado una expansión sostenida que consolidó su presencia internacional al alcanzar las 3,831 publicaciones al cierre del periodo. Este avance, sumado al posicionamiento de la India en el tercer lugar mundial, evidencia una reconfiguración del mapa científico: la producción ya no se concentra únicamente en las potencias tradicionales, sino que se ha diversificado hacia economías emergentes que ahora operan como nuevos centros de innovación.

Figura 4

Evolución de la producción científica anual por país en el área de salud (Figura propia con datos de SCOPUS)



Discusión

El análisis integral de los datos muestra que la producción científica en salud atraviesa una transformación estructural profunda, impulsada por la convergencia entre tecnologías disruptivas y cambios radicales en los modelos editoriales contemporáneos. A diferencia del impacto puntual generado por la publicación del Genoma Humano en 2003, la IAG, especialmente en el periodo 2024-2025, opera como una herramienta transversal cuya capacidad de optimización explica la aceleración simultánea en ámbitos tan diversos como la gestión de procesos (Shonubi, 2026), el diagnóstico clínico (Chen et al., 2024), la innovación farmacológica (Yu et al., 2024) y el modelado de simulaciones complejas (Karmakar et al., 2025). Este fenómeno ha permitido superar contracciones previas en el volumen de publicaciones y ha impulsado un cambio de paradigma orientado hacia la adaptabilidad y la personalización del aprendizaje científico (Almansour & Alfhaid, 2024; Liu et al., 2023; Chen et al., 2024; Nana & Marshall, 2025).

En el plano editorial, esta metamorfosis tecnológica ha dado lugar a un nuevo paradigma en el que la expansión de revistas de acceso abierto y gran escala, como *Scientific Reports* y *PLOS One*, marca el fin de la era de la *escasez editorial* (Nana & Marshall, 2025). El liderazgo de estas megarrevistas evidencia que el valor de la difusión científica

se desplaza hacia la velocidad, el acceso abierto y el rigor técnico, por encima del antiguo prestigio anclado en la exclusividad de la cabecera. Sin embargo, el volumen masivo de publicaciones introduce tensiones significativas en la economía de la atención científica. El modelo tradicional de filtrado previo ha sido sustituido por una lógica de publicación masiva asistida por algoritmos, que obliga a los investigadores a desarrollar estrategias de curaduría algorítmica para navegar entre millones de artículos anuales. Esta saturación, descrita como *ruido epistémico*, incrementa el riesgo de fragmentar el conocimiento en silos digitales de difícil validación humana, donde la disponibilidad de información supera con creces la capacidad de síntesis de la comunidad médica global (Maupin et al., 2025). En consecuencia, la inmediatez en el acceso a los datos debe equilibrarse con marcos de gobernanza que aseguren la trazabilidad y la reproducibilidad de los hallazgos generados y gestionados por redes neuronales, consolidando un ecosistema en el que la eficiencia operativa de la revista resulta tan determinante como el contenido mismo de la investigación (Nana & Marshall, 2025).

De forma paralela a la reconfiguración editorial, la descentralización del conocimiento ha provocado un giro geopolítico sin precedentes en la ciencia global. Se observa un desplazamiento irreversible de las potencias tradicionales de Occidente frente al ascenso meteórico de China e India (Yu et al., 2024). Este cambio no es únicamente cuantitativo, sino que expresa una apuesta estratégica por la infraestructura digital y la formación de talento humano en economías emergentes. China ha consolidado su liderazgo al integrar la IAG desde las etapas experimentales, reduciendo de forma drástica el ciclo que va del laboratorio a la publicación final (Xie et al., 2025). En este escenario multipolar, el caso de México es especialmente relevante: aunque su volumen total es menor, el crecimiento sostenido hacia unas 3,800 publicaciones anuales estimadas para 2025 indica una inserción progresiva y exitosa en las redes globales de innovación. No obstante, esta descentralización del conocimiento reabre debates sobre la soberanía tecnológica. Cuando la producción científica de países emergentes depende de infraestructuras de IA controladas por corporaciones transnacionales del Norte Global, emerge el riesgo de un neocolonialismo de datos, en el que la validación del conocimiento propio queda mediada por algoritmos con sesgos exógenos (Adarkwah et al., 2025).

Un eje de debate central en esta nueva ecología del conocimiento es la emergencia de la llamada *Ciencia Sintética*. La capacidad de los modelos generativos para producir, a gran escala, informes de casos clínicos y síntesis de evidencia tensiona los parámetros

clásicos de integridad académica (Shonubi, 2026). El problema supera el plagio convencional y se sitúa en el terreno de la *alucinación científica*: la generación de datos que, aunque estadísticamente coherentes, carecen de una base empírica verificable (Hu et al., 2025). En este contexto, la arquitectura editorial del futuro deja de ser un soporte pasivo para convertirse en un motor activo que utiliza la IA para curar, difundir y anticipar tendencias de investigación. La democratización del acceso, impulsada por el modelo de acceso abierto, constituye un avance innegable en materia de equidad global; sin embargo, dicha apertura debe articularse con una gobernanza ética robusta que impida la deshumanización del arbitraje científico (Bouderhem, 2024).

La ciencia en salud se configura hoy como una singularidad bibliométrica donde la intersección entre IAG e infraestructura digital ha dado lugar a un ecosistema saturado que exige vigilancia sostenida. El desafío para las instituciones académicas y los sistemas de salud consiste en garantizar que el incremento exponencial de la productividad se traduzca en mejoras reales y tangibles en el bienestar humano, evitando que la producción científica se reduzca a una maquinaria de generación masiva de papel digital desprovista de impacto social (Bouderhem, 2024; Hu et al., 2025). En este horizonte, la curiosidad, el rigor y el juicio crítico humanos deben mantenerse como pilares irrenunciables, utilizando la tecnología como motor transversal para enfrentar crisis sanitarias globales, al tiempo que se salvaguarda la autonomía intelectual del investigador frente al avance de la automatización generativa.

Conclusiones

La transformación estructural de la ciencia en salud, impulsada por la IAG y los nuevos modelos editoriales, configura un punto de inflexión histórico que redefine, tanto la forma de producir conocimiento como los centros desde donde se ejerce el liderazgo científico. La convergencia entre tecnologías disruptivas y megarrevistas de acceso abierto ha permitido una expansión sin precedentes del volumen de publicaciones, optimizando procesos de diagnóstico, innovación farmacológica y modelado complejo, pero al costo de generar un *ruido epistémico* que pone en tensión la capacidad de filtrado y síntesis de la comunidad médica global. En paralelo, la emergencia de una ciencia multipolar, con el ascenso de China, India y el crecimiento sostenido de países como México, demuestra que la capacidad de articular infraestructura digital propia, formar talento y asegurar soberanía tecnológica será decisiva para evitar nuevas formas de neocolonialismo de datos donde los algoritmos externos determinen la validación del conocimiento local.

En este escenario, la irrupción de la *Ciencia Sintética* y el uso masivo de modelos generativos para producir casos clínicos, revisiones y datos simulados obligan a replantear de raíz la ética de la publicación y los mecanismos de verificación, incorporando trazabilidad robusta, transparencia en el grado de intervención algorítmica y marcos de gobernanza que prevengan alucinaciones científicas y sesgos estructurales. La democratización del acceso, aunque innegablemente valiosa, ya no puede analizarse separada de la necesidad de nuevas competencias críticas en investigadores y profesionales de la salud, capaces de ejercer una curaduría algorítmica informada y de mantener la autonomía intelectual frente a la automatización. En última instancia, el éxito de esta metamorfosis dependerá de que las instituciones científicas logren equilibrar eficiencia y vigilancia ética: solo si la IA se integra como colaborador responsable, y no como sustituto del criterio humano, la hiper-productividad bibliométrica podrá traducirse en mejoras reales, equitativas y sostenibles para la salud global, evitando que la ciencia quede reducida a una fábrica de papel digital sin impacto social efectivo.

Referencias

Adarkwah, M. A., Badu, S. A., Osei, E. A., Adu-Gyamfi, E., Odame, J., & Schneider, K. (2025). ChatGPT in healthcare education: A double-edged sword of trends, challenges, and opportunities. *Discover Education*, *4*(1), 14. https://doi.org/10.1007/s44217-024-00393-3

Almansour, M., & Alfhaid, F. M. (2024). Generative artificial intelligence and the personalization of health professional education: A narrative review. *Medicine*, *103*(31), e38955. https://doi.org/10.1097/MD.0000000000038955

Bouderhem, R. (2024). Shaping the future of AI in healthcare through ethics and governance. *Humanities and Social Sciences Communications*, *11*(1), 416. https://doi.org/10.1057/s41599-024-02894-w

Chen, A., Liu, L., & Zhu, T. (2024). Advancing the democratization of generative artificial intelligence in healthcare: A narrative review. *Journal of Hospital Management and Health Policy*, 8, 12-12. https://doi.org/10.21037/jhmhp-24-54

De Vere Hunt, I. J., Jin, K.-X., & Linos, E. (2025). A framework for considering the use of generative AI for health. *Npj Digital Medicine*, *8*(1), 297. https://doi.org/10.1038/s41746-025-01695-y

Hu, N., Jiang, X. Q., Wang, Y. D., Kang, Y. M., Xia, Z., Chen, H. H., Duan, S. N., & Chen, D. X. (2025). Status and perceptions of ChatGPT utilization among medical students: A survey-based study. *BMC Medical Education*, *25*(1), 831. https://doi.org/10.1186/s12909-025-07438-7

Karmakar, A., Shaw, A., Rakshit, S., Chakraborty, S., Biswas, Sitanath, Sahoo, S., & Biswas, Suparna. (2025). The role of generative AI in medical image synthesis: A review. *Discover Applied Sciences*, *7*(10), 1219. https://doi.org/10.1007/s42452-025-07714-7

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-Related research and perspective towards the future of large language models. *Meta-Radiology*, *1*(2), 100017. https://doi.org/10.1016/j.metrad.2023.100017

Maupin, D., Suchak, T., Barnett, A., & Spick, M. (2025). Dramatic increases in redundant publications in the Generative AI era. *BMC Medicine*, *24*(1), 29. https://doi.org/10.1186/s12916-025-04569-y

Mirzaei, T., Amini, L., & Esmaeilzadeh, P. (2024). Clinician voices on ethics of LLM integration in healthcare: A thematic analysis of ethical concerns and implications. *BMC Medical Informatics and Decision Making*, *24*(1), 250. https://doi.org/10.1186/s12911-024-02656-3

Nana, V. K., & Marshall, M. T. (2025). Generative Artificial Intelligence in Healthcare: A Bibliometric Analysis and Review of Potential Applications and Challenges. *AI*, *6*(11), 278. https://doi.org/10.3390/ai6110278

Raman, R., Kumar Nair, V., Nedungadi, P., Kumar Sahu, A., Kowalski, R., Ramanathan, S., & Achuthan, K. (2024). Fake news research trends, linkages to generative artificial intelligence and sustainable development goals. *Heliyon*, *10*(3), e24727. https://doi.org/10.1016/j.heliyon.2024.e24727

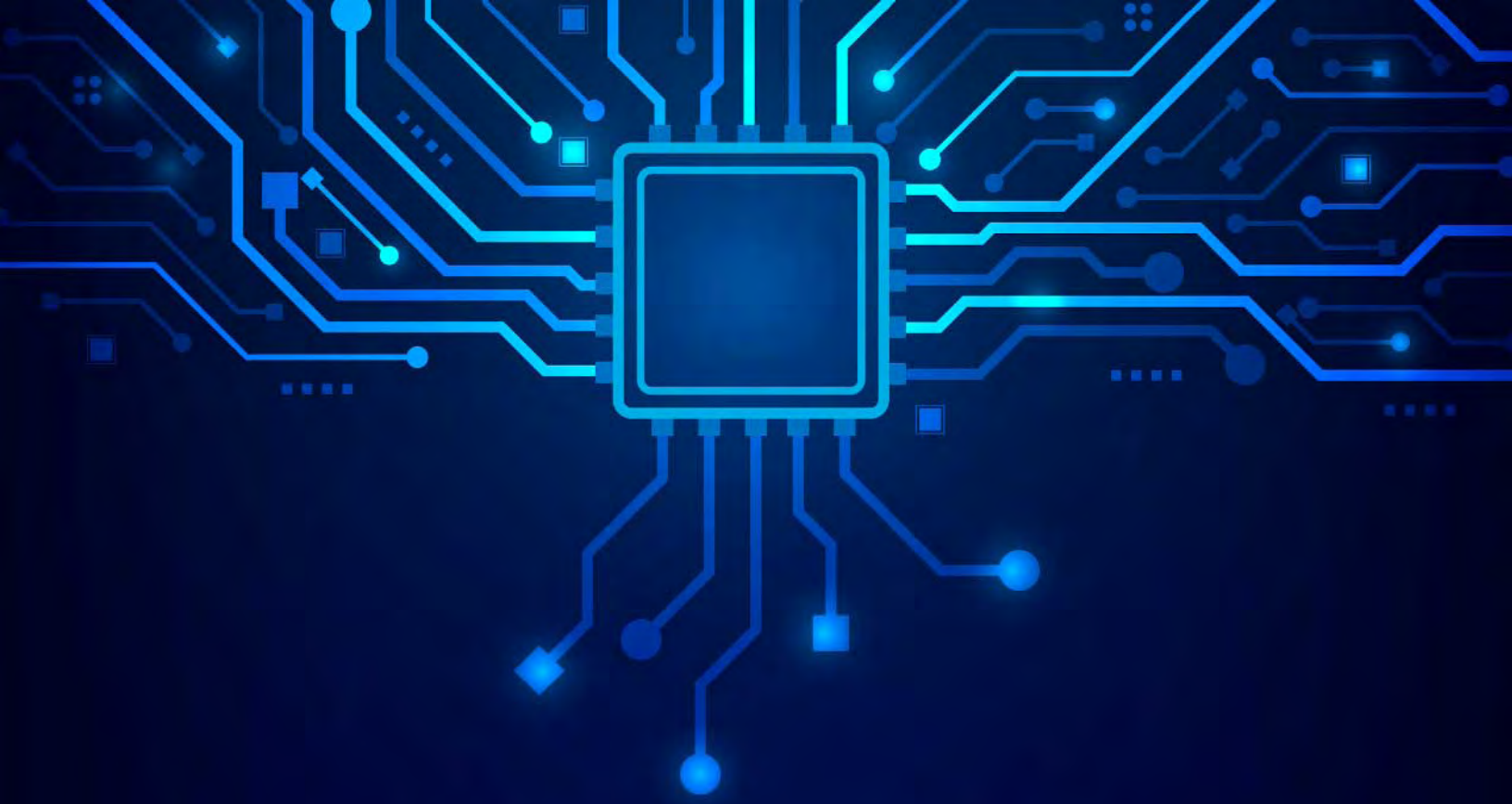
Reddy, S. (2024). Generative AI in healthcare: An implementation science informed translational path on application, integration and governance. *Implementation Science*, *19*(1), 27. https://doi.org/10.1186/s13012-024-01357-9

Shonubi, O. A. (2026). Unwrapping generative AI paradigms for product and service innovation differentiation. *Journal of Strategy & Innovation*, 37(1), 200562. https://doi.org/10.1016/j.jsinno.2025.200562

Xie, Y., Zhai, Y., & Lu, G. (2025). Evolution of artificial intelligence in healthcare: A 30-year bibliometric study. *Frontiers in Medicine*, 11, 1505692. https://doi.org/10.3389/fmed.2024.1505692

Yu, H., Fan, L., Li, L., Zhou, J., Ma, Z., Xian, L., Hua, W., He, S., Jin, M., Zhang, Y., Gandhi, A., & Ma, X. (2024). Large Language Models in Biomedical and Health Informatics: A Review with Bibliometric Analysis. *Journal of Healthcare Informatics Research*, *8*(4), 658-711. https://doi.org/10.1007/s41666-024-00171-8

Zaghir, J., Naguib, M., Bjelogrlic, M., Névéol, A., Tannier, X., & Lovis, C. (2024). Prompt Engineering Paradigms for Medical Applications: Scoping Review. *Journal of Medical Internet Research*, 26, e60501. https://doi.org/10.2196/60501



15.

Factores críticos de éxito para implementar la inteligencia artificial en las unidades de cuidados intensivos

Carlos Zepeda-Lugo
Universidad Autónoma de Baja California, México
ORCID: 0000-0002-9444-3106

Marcos Sanchez-Lizarraga
Universidad Autónoma de Sinaloa, México
ORCID: 0000-0001-8730-9913

Ivette Selene Marañón Lizárraga
Universidad Autónoma de Sinaloa, México
ORCID: 0009-0003-5882-757X

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.15

15.

Factores críticos de éxito para implementar la inteligencia artificial en las unidades de cuidados intensivos

Introducción

Las Unidades de Cuidados Intensivos (UCI) atienden a pacientes críticamente enfermos que requieren atención especializada (Pérez Solórzano et al., 2024). Su propósito es salvar la vida del paciente crítico y mantener su estabilidad y calidad de vida (Muñoz et al., 2025; Pérez & Ordóñez, 2024). Actualmente, las UCI son un cuello de botella en el flujo hospitalario debido a su capacidad limitada y a la creciente demanda, a lo que se suma el impacto de las Infecciones Asociadas a la Atención Sanitaria (IAAS) (Akbari-Moghaddam et al., 2025; Quiroz et al., 2026). Estas infecciones, en su mayoría bacterianas, son causa frecuente de ingreso a la UCI y se relacionan con una mayor Duración de la Estancia Hospitalaria (DEH) (Agüero Milanés et al., 2021; Bagshaw et al., 2020; Bereket et al., 2012; Xia et al., 2016). La pandemia por SARS-CoV-2 intensificó las presiones y aceleró la reorganización hospitalaria (Martínez Muñoz et al., 2025).

En este contexto, la inteligencia artificial (IA) se consolidó para optimizar recursos y apoyar decisiones clínicas, incluida la predicción de la DEH y la preparación del alta del paciente de la UCI (Alsinglawi et al., 2024; Bishop et al., 2021; Schwalbe & Wahl, 2020; Topol, 2019; Workum et al., 2026; Yepes-Barreto et al., 2026). No obstante, el rendimiento predictivo no asegura impacto clínico, ya que persisten barreras sociotécnicas y la literatura aún trata de forma limitada los Factores Críticos de Éxito (FCE) para implementar modelos de IA (Kalimouttou et al., 2026; Merhi, 2021; Nair et al., 2024). Por lo tanto, el objetivo de la presente investigación fue identificar los FCE que se relacionan con la implementación exitosa de modelos de IA para la predicción de DEH en UCI.

Factores críticos de éxito en la implementación de inteligencia artificial

La implementación exitosa de la IA en las UCI depende de FCE como una gobernanza robusta, que aborda regulaciones éticas y legales para facilitar la integración en los flujos de trabajo hospitalarios (Hassan et al., 2024; Marwood, 2025; Rahimi et al., 2024). En este sentido, la confianza emerge como un catalizador principal de la adopción de la IA, aunque se ve afectada por barreras como la falta de transparencia en los algoritmos, lo que requiere marcos legislativos que la fortalezcan (Hassan et al., 2024; Scipion et al., 2025). A su vez, estos elementos también mejoran la calidad general de la atención al paciente y promueven una mayor familiaridad con la tecnología entre los profesionales de la salud (D'Hondt et al., 2022; Eltawil et al., 2023). De este modo, se supera la percepción de amenaza a la autonomía profesional, que es un aspecto común en entornos médicos (Eltawil et al., 2023).

En el sentido de las UCI, las soluciones prácticas para afrontar los cambios de los datos clínicos a lo largo del tiempo y en el rendimiento de los modelos predictivos, se convierten en factores esenciales para permitir que los algoritmos de la IA se adapten a variaciones en las características de los pacientes disponibles (D'Hondt et al., 2022; Rahimi et al., 2024). Además, la escasez de datos se mitiga mediante implementaciones reutilizables que fomentan la repetibilidad científica y la transferencia efectiva de la práctica clínica (D'Hondt et al., 2022; Hassan et al., 2024).

Por otra parte, la participación del personal de la salud contribuye a reducir la resistencia inicial al cambio, así como en abordar de manera explícita las consideraciones legales y éticas asociadas al uso de estos sistemas, lo que favorece una adopción responsable (Hassan et al., 2024; Marwood, 2025; Scipion et al., 2025). Igualmente, factores como las dinámicas relacionales en equipos multidisciplinarios y las características técnicas de los sistemas, cuyo impacto depende del contexto de la implementación, deben interpretarse cuidadosamente para maximizar su efecto positivo (D'Hondt et al., 2022; Rahimi et al., 2024; Scipion et al., 2025).

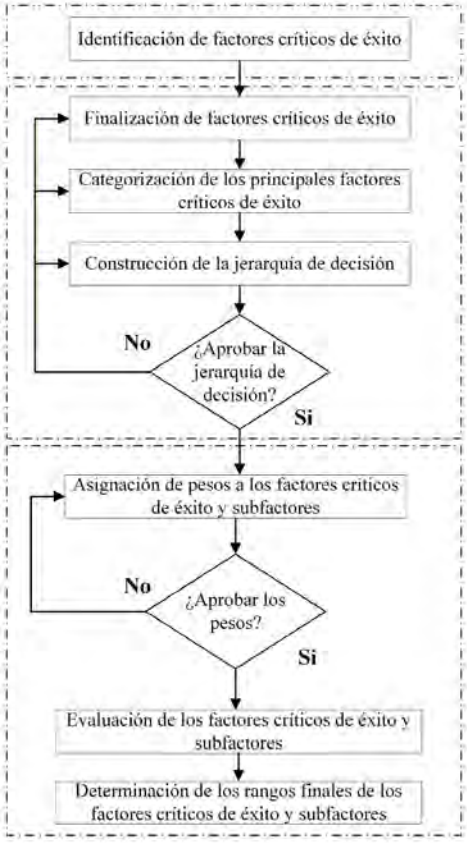
Finalmente, los marcos éticos garantizan transparencia y equidad en el acceso, interpretados como facilitadores clave para eliminar obstáculos en entornos de alto riesgo como las UCI (Hassan et al., 2024; Marwood, 2025; Scipion et al., 2025). Estos factores críticos guían una adopción sostenible y efectiva de la IA en cuidados intensivos si se desarrollan en conjunto.

Método de investigación

Primero, se identificaron los FCE mediante la revisión de artículos científicos de *PubMed*, *Web of Science*, *ScienceDirect*, *SciELO*, *IEEE Xplore* y *Google Scholar*, incluyendo estudios sobre implementación de IA en una UCI (Tabla 1). Posteriormente, quince expertos en UCI, gestión hospitalaria, IA aplicada en salud y sistemas de información validaron los factores encontrados en la literatura, más la adición de factores complementarios, quienes definieron 18 factores organizados en tres categorías: Factores Organizacionales, Factores Psicológicos-Humanos y Factores Técnicos. Para evaluar y estimar la importancia de los FCE, se utilizó la metodología del Proceso Analítico Jerárquico (PAJ) (Figura 1) (Altawaiha et al., 2025; Merhi, 2021). Este permitió estructurar jerarquías, realizar comparaciones pareadas y calcular pesos relativos para priorizar los factores en decisiones multicriterio sin la necesidad de contar con un tamaño de muestra estadísticamente significativo (Truong & Le, 2024; Doulabi et al., 2024).

Figura 1

Proceso metodológico de investigación



Nota. Adaptado de Merhi (2021).

Para ponderar los factores, cada experto comparó todos los pares posibles siguiendo el método de comparaciones pareadas, donde cada experto evalúa la importancia relativa entre dos factores utilizando la escala de nueve puntos; es decir, cuando un factor *i* tiene mayor importancia que un factor *j*, se asignan valores de 1 a 9 según la intensidad; por otra parte, cuando *j* es más importante que *i*, se utiliza el valor recíproco correspondiente (Tabla 1).

Tabla 1
Escala de medición

Comparación entre *i* y *j*	a *i j*	1/a *i j*
Igual importancia	1	1/1
Importancia ligera de uno sobre otro	3	1/3
Importancia fuerte	5	1/5
Importancia muy fuerte	7	1/7
Importancia extrema	9	1/9

Con estas comparaciones se construyeron matrices de comparaciones pareadas y se calcularon los pesos de prioridad (importancia relativa de cada factor), el Índice de Consistencia (IC), que mide la coherencia de los juicios, y la Razón de Consistencia (RC), que normaliza el valor de IC. La RC detecta inconsistencias en los juicios de los expertos, como transitividades violadas entre preferencias. Se establecen criterio de aceptación de la RC: valores entre 0 y 0.1 para matrices de tamaño mayores que 4×4, ya que valores superiores indican inconsistencia inaceptable que requiere repetir las comparaciones (Haq & Kannan, 2006; Saaty, 2001).

En este caso, se utilizó una matriz de 18×18 lo que define un índice de consistencia aleatorio RI=1.59, que según la Tabla 2, es el índice utilizado para matrices iguales o mayores a 15. Con estos valores se calcula la RC, donde se divide el IC entre el índice de consistencia aleatoria. Finalmente, se realizó la síntesis PAJ para estimar pesos locales por categoría y pesos globales respecto al objetivo general, así como los pesos de *ranking* final por factor para identificar las prioridades de una exitosa implementación de modelos de IA en una UCI.

Tabla 2
Índice de consistencia aleatoria

Tamaño de la matriz	Índice de consistencia aleatoria
1	0
2	0
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49
11	1.51
12	1.48
13	1.56
14	1.57
15	1.59

Nota. Tomada de Saaty (2001).

Resultados

La caracterización demográfica de los 15 expertos fue mayormente del género masculino con un 73.3% y un rango de edad de 30-50 años con un 60%. En cuanto al nivel educativo, el nivel de maestría predominó con un 53.3% y el mayor puesto laboral mencionado fue el de médico especialista con un 46.7%. Por último, el 60% de los expertos procedían del hospital público general.

En cuanto a los FCE analizados por los expertos, se categorizaron 18 factores en tres dimensiones (Figura 2). La primera categoría de Factores Organizacionales cuenta con 4 factores; seguido por los Factores Psicológicos y Humanos con 6 factores; y, por último, la categoría de Factores Técnicos con 8 factores.

Figura 2

Factores críticos para implementar inteligencia artificial en cuidados intensivos

Por otra parte, en la Tabla 3 se presentan los resultados de la priorización de FCE obtenidos mediante la metodología PAJ, donde se observa que la categoría de Factores Técnicos concentró el mayor peso porcentual (0.51), seguidos por los Factores Psicológicos y Humanos (0.30) y por último los Factores Organizacionales (0.19). En cuanto los factores, el mayor peso global correspondieron a “Confianza y explicabilidad” (0.0876; prioridad global 1), seguido por “Calidad y disponibilidad de datos” (0.08144; prioridad global 2) y “Robustez y generalización” (0.07821; prioridad global 3). En contraste, los valores más bajos se observaron en “Miedo al reemplazo laboral” (0.02827; prioridad global 17) y “Latencia y tiempo de respuesta” (0.02695; prioridad global 18).

Tabla 3

Pesos de prioridad en el árbol de decisiones PAJ

Categoría-dimensión	Peso porcentual entre las categorías	Peso porcentual dentro de las categorías	Orden de prioridad entre factores	Peso porcentual entre los factores	Orden de prioridad entre categorías
Factores Organizacionales	**0.19**				
Alineación estratégica		0.14187	4	0.03597	15
Integración en el flujo de trabajo		0.36821	1	0.0748	4
Gobernanza de datos		0.30794	2	0.06172	9

Tabla 3

Pesos de prioridad en el árbol de decisiones PAJ

Categoría-dimensión	Peso porcentual entre las categorías	Peso porcentual dentro de las categorías	Orden de prioridad entre factores	Peso porcentual entre los factores	Orden de prioridad entre categorías
Cumplimiento normativo y legal		0.18198	3	0.05017	11
Factores Psicológicos y Humanos	**0.30**				
Confianza y explicabilidad		0.28463	1	0.0876	1
Miedo al reemplazo laboral		0.08241	6	0.02827	17
Participación en el co-diseño		0.14178	4	0.03959	13
Usabilidad y experiencia de usuario		0.17392	3	0.06463	7
Gestión de expectativas		0.10953	5	0.034	16
Ética y percepción de sesgo		0.20773	2	0.07021	5
Factores Técnicos	**0.51**				
Calidad y disponibilidad de datos		0.20366	1	0.08144	2
Interoperabilidad		0.12184	4	0.06197	8
Estandarización (FHIR/HL7)		0.09077	6	0.04933	12
Robustez y generalización		0.15543	2	0.07821	3
Seguridad y privacidad		0.14625	3	0.0651	6
Escalabilidad y MLOps		0.10981	5	0.05324	10
Latencia y tiempo de respuesta		0.08359	8	0.02695	18
Anonimización y desidentificación		0.08865	7	0.0368	14

Discusión

La dimensión de Factores Técnicos tiene la mayor prioridad para implementar IA en las UCI con un valor de 0.51, seguidos por Psicológicos y Humanos con 0.30 y los Organizacionales con 0.19. Esto indica que primero se requiere solidez técnica en datos y modelos de IA para lograr una aceptación humana y obtener soporte institucional (Alsallal et al., 2025; Charan et al., 2023; Greco et al., 2021; Janiesch et al., 2021; Mandala et al., 2023). Esta jerarquía concuerda con la evidencia que destaca calidad y disponibilidad de datos, robustez y generalización del modelo e integración al flujo de trabajo como habilitadores del impacto clínico más allá del rendimiento predictivo aislado (D'Hondt et al., 2022; Hassan et al., 2024; Kalimouttou et al., 2026; Merhi, 2021; Rahimi et al., 2024; Scipion et al., 2025).

La alta prioridad de calidad/disponibilidad de datos y de robustez/generalización indica que se necesitan datos completos, estandarizados y longitudinales para entrenar modelos confiables y transferibles entre hospitales. Una barrera clave para llevar modelos de IA clínica como MIMIC-III o eICU a contextos locales (Erdogan Yildirim & Canayaz, 2024; Fan et al., 2025; Hu et al., 2024; Junior et al., 2024; Ramakrishnaiah et al., 2025; Tella & Balasundaram, 2025). Además, la evidencia indica que la estandarización (FHIR/HL7), la interoperabilidad y las prácticas de operación de modelos son necesarias para mantener el desempeño en uso real, actualizar los modelos y gestionar la deriva de datos coherente de acuerdo con el orden de prioridad obtenida (Altawaiha et al., 2025; D'Hondt et al., 2022; Merhi, 2021; Truong & Le, 2024; Doulabi et al., 2024).

Por otra parte, la seguridad y privacidad, junto con la anonimización y desidentificación, quedan en posiciones media-altas, coherentes con marcos ético-legales que exigen proteger datos sensibles como condición para desplegar IA en UCI (Hassan et al., 2024; Marwood, 2025; Rahimi et al., 2024; Scipion et al., 2025). Además, el uso secundario de datos, el riesgo de reidentificación y la gobernanza poco clara frenan proyectos incluso técnicamente maduros, lo que explica el buen posicionamiento de la gobernanza de datos entre los factores organizacionales (Hassan et al., 2024; Marwood, 2025; Rahimi et al., 2024; Saaty, 1987, 2001).

En lo Psicológico y Humano, que la confianza y explicabilidad tenga el mayor peso global, incluso sobre la calidad de datos, concuerda con evidencia cualitativa la cual define que el persona clínico adoptan modelos de IA cuando son comprensibles, auditables y coherentes con su razonamiento y no solo precisos (Chen et al., 2024; D'Hondt et al., 2022;

Eltawil et al., 2023; Hassan et al., 2024; Nair et al., 2024; Scipion et al., 2025). En entornos de alto riesgo, la opacidad aumenta percepciones de sesgo y reduce el uso, por lo que ética y percepción de sesgo también queda entre los factores mejor clasificados (Chen et al., 2024; Eltawil et al., 2023; Hassan et al., 2024; Marwood, 2025; Rahimi et al., 2024; Scipion et al., 2025).

La prioridad de usabilidad/experiencia de usuario y del co-diseño coincide con la evidencia sobre las interfaces mal alineadas al flujo de una UCI la cual aumentan la carga cognitiva y resistencia, mientras el diseño participativo mejora aceptación y apropiación local (D'Hondt et al., 2022; Hassan et al., 2024; Rahimi et al., 2024; Scipion et al., 2025). En cambio, el bajo peso del miedo al reemplazo laboral sugiere que pesan más preocupaciones sobre calidad, responsabilidad y control clínico, coherente con la IA en una UCI como apoyo a la decisión y no sustitución (Eltawil et al., 2023; Hassan et al., 2024; Nair et al., 2024; Schwalbe & Wahl, 2020; Topol, 2019; Yepes-Barreto et al., 2026).

En lo Organizacional, que la integración en el flujo de trabajo supere a la alineación estratégica y al cumplimiento normativo sugiere que las barreras inmediatas están por adaptar la IA en procesos cotidianos más que en políticas *en papel* (Bagshaw et al., 2020; Bishop et al., 2021; Hassan et al., 2024; Janssens et al., 2025; Kalimouttou et al., 2026; Rahimi et al., 2024). Incluso con apoyo directivo y aval ético, la falta de integración con sistemas clínicos y la disrupción de rutinas puede impedir el uso diario, lo que explica la menor posición de la alineación estratégica (Hassan et al., 2024; Merhi, 2021; Nair et al., 2024; Rahimi et al., 2024).

Un resultado importante es la baja prioridad otorgada a latencia y tiempo de respuesta, pese a que las UCI dependen de decisiones sensibles al tiempo. Esto sugiere, por un lado, que la mayoría de modelos para predicción de DEH operan en escalas de horas días y no exigen decisiones en segundos o minutos; y por otro lado, los tiempos de cómputo moderados son aceptables (D'Hondt et al., 2022; Fan et al., 2025; Rahimi et al., 2024; Ramakrishnaiah et al., 2025; Tella & Balasundaram, 2025).

El presente estudio tiene limitaciones porque el uso del PAJ depende de un panel pequeño y puede reflejar sesgos del contexto institucional, la familiaridad con la IA y las percepciones de riesgos y beneficios; por lo tanto, los pesos podrían no generalizarse a otras UCI o sistemas de salud (Altawaiha et al., 2025; Haq & Kannan, 2006; Merhi, 2021; Saaty, 2001; Truong & Le, 2024; Doulabi et al., 2024). Además, los FCE se identificaron mediante una revisión exploratoria que podría omitir factores emergentes o propios de entornos de

bajos recursos, y la jerarquía aún no se validó de forma prospectiva en implementaciones reales, de modo que su capacidad para predecir éxito operativo y clínico sigue siendo una hipótesis de carácter exploratoria (Agüero Milanés et al., 2021; Bagshaw et al., 2020; Janssens et al., 2025; Kalimouttou et al., 2026; Merhi, 2021).

Los hallazgos aportan valor teórico y práctico que, en teoría, la estructura tripartita de Factores Técnicos, Factores Psicológicos y Humanos y Factores Organizacionales, con su priorización cuantitativa, cubre un vacío sobre los FCE para implementar modelos de IA en las UCI ofreciendo un marco integrador que conecta gobernanza de datos, confianza clínica e integración operativa (Hassan et al., 2024; Merhi, 2021; Rahimi et al., 2024; Schwalbe & Wahl, 2020; Scipion et al., 2025; Topol, 2019). En la práctica, la jerarquía orienta a focalizar recursos en datos de alta calidad, modelos robustos y explicables, interoperabilidad e integración al flujo de trabajo, apoyados por formación, co-diseño y marcos éticos, son necesarios para escalar la IA en la predicción DEH y optimizar recursos en las UCI (Bagshaw et al., 2020; Bishop et al., 2021; D'Hondt et al., 2022; Hassan et al., 2024; Kalimouttou et al., 2026; Rahimi et al., 2024).

Conclusión

Este estudio prioriza los Factores Críticos de Éxito para implementar modelos de IA predictivos en la duración de la estancia hospitalaria en unidades de cuidados intensivos, combinando una revisión bibliográfica y aplicando un juicio de experto mediante el Proceso Analítico Jerárquico. En cuanto los Factores Técnicos, Factores Psicológicos y Humanos y los Factores Organizacionales, estas categorías que integran los 18 FCE, destacan que el éxito depende primero de datos y modelos sólidos, seguido de una confianza clínica y por último, el poseer un soporte institucional.

La integración de evidencia científica con expertos y el uso de PAJ ofrecen un marco práctico para planificar recursos, aunque la revisión exploratoria, el panel limitado y la ausencia de validación prospectiva restringen la generalización. Se recomienda replicar el marco en otros contextos, comparar prioridades por tipo de UCI y recursos, y vincular los factores principales a resultados objetivos como menor estancia prolongada, mejor flujo hospitalario y mayor satisfacción. La jerarquía propuesta cierra la brecha entre modelos de alto rendimiento y su adopción efectiva, proporcionando una hoja de ruta para decisiones técnicas, humanas y organizacionales en IA para las UCI.

Referencias

Agüero Milanés, A. M., Infante Rondón, K. Z., & Delgado Llorca, F. E. (2021). Infecciones nosocomiales por bacterias gram negativas y estadía prolongada en cuidados intensivos pediátricos. *Revista Habanera de Ciencias Médicas* (Infomed), *20*(3), e3608.

Akbari-Moghaddam, M., Down, D. G., Li, N., Eastwood, C., Mehrem, A. A., & Howlett, A. (2025). Data-Driven Bed Occupancy Planning in Intensive Care Units Using Mt/Gt/ Queueing Models (arXiv:2510.02852). *arXiv*. https://doi.org/10.48550/arXiv.2510.02852

Alsallal, M., Habeeb, M. S., Vaghela, K., Malathi, H., Vashisht, A., Sahu, P. K., Singh, D., Al-Hussainy, A. F., Aljanaby, I. A., Sameer, H. N., Athab, Z. H., Adil, M., Yaseen, A., & Farhood, B. (2025). Artificial intelligence in gastric cancer: A systematic review of machine learning and deep learning applications. *Abdominal Radiology*. https://doi.org/10.1007/s00261-025-05181-7

Alsinglawi, B. S., Alnajjar, F., Alorjani, M. S., Al-Shari, O. M., Munoz, M. N., & Mubin, O. (2024). Predicting Hospital Stay Length Using Explainable Machine Learning. *IEEE Access*, 12, 90571–90585. https://doi.org/10.1109/ACCESS.2024.3421295

Altawaiha, I., Atan, R., Yaakob, R. B., & Abdullah, R. B. H. (2025). Assessing and prioritizing crucial drivers for CloudIoT-based healthcare adoption: An analytic hierarchy process approach. *International Journal of Information Technology, 17*(8), 5061–5078. https://doi.org/10.1007/s41870-024-01742-z

Bagshaw, S. M., Tran, D. T., Opgenorth, D., Wang, X., Zuege, D. J., Ingolfsson, A., Stelfox, H. T., & Thanh, N. X. (2020). Assessment of Costs of Avoidable Delays in Intensive Care Unit Discharge. *JAMA Network Open*, *3*(8), e2013913. https://doi.org/10.1001/jamanetworkopen.2020.13913

Bereket, W., Hemalatha, K., Getenet, B., Wondwossen, T., Solomon, A., Zeynudin, A., & Kannan, S. (2012). Update on Bacterial Nosocomial Infections. *European Review for Medical and Pharmacological Sciences*, *16*(8), 1039–1044.

Bishop, J. A., Javed, H. A., el-Bouri, R., Zhu, T., Taylor, T., Peto, T., Watkinson, P., Eyre, D. W., & Clifton, D. A. (2021). Improving patient flow during infectious disease outbreaks using machine learning for real-time prediction of patient readiness for discharge. *PLOS ONE*, *16*(11), e0260476. https://doi.org/10.1371/journal.pone.0260476

Charan, G. S., Charan, A. S., Khurana, M. S., & Narang, G. S. (2023). Impact of Analytics Applying Artificial Intelligence and Machine Learning on Enhancing Intensive Care Unit: A Narrative Review. *Galician Medical Journal*, *30*(4). https://doi.org/10.21802/e-GMJ2023-A06

Chen, F. K., Heath Jeffery, R. C., & Fung, A. T. (2024). Bringing Mohammad to the mountain: New strategies for intravitreal therapy service delivery. *Clinical & Experimental Ophthalmology*, *52*(8), 797–799. https://doi.org/10.1111/ceo.14457

D'Hondt, E., Ashby, T. J., Chakroun, I., Koninckx, T., & Wuyts, R. (2022). Identifying and evaluating barriers for the implementation of machine learning in the intensive care unit. *Communications Medicine*, *2*(1), 162. https://doi.org/10.1038/s43856-022-00225-1

Doulabi, R. Z., Asnaashari, E., Shaygan, D. S., & Amirkardoost, A. (2024). The identification and prioritization of critical success factors in healthcare projects through the application of the Analytic Hierarchy Process (AHP). *Power System Technology*, *48*(1), 1643-1653.

Eltawil, F. A., Atalla, M., Boulos, E., Amirabadi, A., & Tyrrell, P. N. (2023). Analyzing Barriers and Enablers for the Acceptance of Artificial Intelligence Innovations into Radiology Practice: A Scoping Review. *Tomography*, *9*(4), 1443–1455. https://doi.org/10.3390/tomography9040115

Erdogan Yildirim, A., & Canayaz, M. (2024). Machine learning-based prediction of length of stay (LoS) in the neonatal intensive care unit using ensemble methods. *Neural Computing and Applications*, *36*(23), 14433–14448. https://doi.org/10.1007/s00521-024-09831-7

Fan, G., Liu, A., & Zhang, C. (2025). ContrastLOS: A Graph-Based Deep Learning Model With Contrastive Pre-Training for Improved ICU Length-of-Stay Prediction. *IEEE Access*, 13, 34132–34148. https://doi.org/10.1109/ACCESS.2025.3540896

Greco, M., Caruso, P. F., & Cecconi, M. (2021). Artificial Intelligence in the Intensive Care Unit. *Seminars in Respiratory and Critical Care Medicine*, *42*(1), 2–9. https://doi.org/10.1055/s-0040-1719037

Haq, A. N., & Kannan, G. (2006). Fuzzy analytical hierarchy process for evaluating and selecting a vendor in a supply chain model. *International Journal of Advanced Manufacturing Technology*, *29*(7–8), 826–835. https://doi.org/10.1007/s00170-005-2562-8

Hassan, M., Kushniruk, A., & Borycki, E. (2024). Barriers to and Facilitators of Artificial Intelligence Adoption in Health Care: Scoping Review. *JMIR Human Factors*, *11*(1), e48633. https://doi.org/10.2196/48633

Hu, T.-L., Chao, C.-M., Wu, C.-C., Chien, T.-N., & Li, C. (2024). Machine Learning-Based Predictions of Mortality and Readmission in Type 2 Diabetes Patients in the ICU. *Applied Sciences*, *14*(18). https://doi.org/10.3390/app14188443

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Janssens, J. M. F., van der Ham, A., Ruwaard, D., & van Merode, F. (2025). Bed Blocking by Hospitalized Patients Awaiting Admission to Intramural Aftercare: A Case Study on Transfer Coordination. *Healthcare*, *13*(16), 2038. https://doi.org/10.3390/healthcare13162038

Junior, J. C., Caneo, L. F., Turquetto, A. L. R., Amato, L. P., Arita, E. C. T. C., Fernandes, A. M. da S., Trindade, E. M., Jatene, F. B., Dossou, P.-E., & Jatene, M. B. (2024). Predictors of in-ICU length of stay among congenital heart defect patients using artificial intelligence model: A pilot study. *Heliyon*, *10*(4). https://doi.org/10.1016/j.heliyon.2024.e25406

Kalimouttou, A., Stevens, R. D., & Pirracchio, R. (2026). Harnessing AI in critical care: Opportunities, challenges and key steps for success. *Thorax*, 8, 183-192. https://doi.org/10.1136/thorax-2024-222125

Mandala, G., Danda, R. R., Nishanth, A., Yasmeen, Z., & Maguluri, K. K. (2023). AI and ML in healthcare: Redefining diagnostics, treatment, and personalized medicine. *International Journal of Applied Engineering & Technology*, *5*(S6), 560–571.

Martínez Muñoz, M., Barba Flores, M. Á., Abadias Medrano, M. J., & Salazar Soler, A. (2025). Transformación de un hospital: De especialidades departamentales a áreas de conocimiento integradas. Un análisis de resultados en salud y experiencia de pacientes y profesionales. *Medicina Clínica*, *165*(5), 107174. https://doi.org/10.1016/j.medcli.2025.107174

Marwood, T. (2025). *Artificial Intelligence (AI) in healthcare: Facilitators and barriers to applying ethics frameworks* [Thesis, Macquarie University]. https://doi.org/10.25949/28585217.v1

Merhi, M. I. (2021). Evaluating the critical success factors of data intelligence implementation in the public sector using analytical hierarchy process. *Technological Forecasting and Social Change*, 173, 121180. https://doi.org/10.1016/j.techfore.2021.121180

Muñoz, L. L. Z., Vera, J. M. V., Remache, A. P. P., & Mejía, B. A. Z. (2025). Gestión de costos hospitalarios con enfoque en UCI: Revisión sistemática. *Conciencia Digital*, *8*(1.1), 100–124. https://doi.org/10.33262/concienciadigital.v8i1.1.3355

Nair, M., Svedberg, P., Larsson, I., & Nygren, J. M. (2024). A comprehensive overview of barriers and strategies for AI implementation in healthcare: Mixed-method design. *PLOS ONE*, *19*(8), e0305949. https://doi.org/10.1371/journal.pone.0305949

Pérez Solórzano, F. A., Zambrano Intriago, G. I., Burbano Vera, M. E., Palma Mejía, V. P., Menéndez Chávez, P. M., & Mendoza Alay, R. E. (2024). Aplicación del paquete BUNDLE ABCDEF en pacientes hospitalizados en la unidad de cuidados intensivos: Revisión Sistemática. *Ciencia Latina: Revista Multidisciplinar*, *8*(4), 1462–1482.

Pérez, J. F., & Ordóñez, J. A. C. (2024). Evaluación de la carga de trabajo de enfermería en UCI usando TISS 28 y NAS: Revisión sistemática. *Revista Conecta Libertad*, 8(4), 1–9.

Quiroz, J. T. G., Crespo, S. E. S., Chavez, K. L. H., & Lemache, S. L. R. (2026). Rol de enfermería en prevención y control de infecciones en UCI asociadas a la atención: Revisión sistemática. *Revista Vive*, *9*(25), 42–60. https://doi.org/10.33996/revistavive.v9i25.455

Rahimi, A. K., Pienaar, O., Ghadimi, M., Canfell, O. J., Pole, J. D., Shrapnel, S., Vegt, A. H. van der, & Sullivan, C. (2024). Implementing AI in Hospitals to Achieve a Learning Health System: Systematic Review of Current Enablers and Barriers. *Journal of Medical Internet Research*, *26*(1), e49655. https://doi.org/10.2196/49655

Ramakrishnaiah, Y., Macesic, N., Webb, G. I., Peleg, A. Y., & Tyagi, S. (2025). EHR-ML: A data-driven framework for designing machine learning applications with electronic health records. *International Journal of Medical Informatics*, 196, 105816. https://doi.org/10.1016/j.ijmedinf.2025.105816

Saaty, R. W. (1987). The analytic hierarchy process—What it is and how it is used. *Mathematical Modelling*, *9*(3), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Saaty, T. L. (2001). Fundamentals of the Analytic Hierarchy Process. En D. L. Schmoldt, J. Kangas, G. A. Mendoza, & M. Pesonen (Eds.), *The Analytic Hierarchy Process in Natural Resource and Environmental Decision Making* (pp. 15–35). Springer Netherlands. https://doi.org/10.1007/978-94-015-9799-9_2

Scipion, C. E. A., Manchester, M. A., Federman, A., Wang, Y., & Arias, J. J. (2025). Barriers to and facilitators of clinician acceptance and use of artificial intelligence in healthcare settings: A scoping review. BMJ Open, 15, e092624. https://doi.org/10.1136/bmjopen-2024-092624

Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *The Lancet*, *395*(10236), 1579–1586. https://doi.org/10.1016/S0140-6736(20)30226-9

Tella, A. K., & Balasundaram, S. R. (2025). Enhanced Prediction of Intensive Care Unit Length of Stay using a Stack Ensemble of Machine Learning Models. *Engineering, Technology & Applied Science Research*, *15*(1), 20367–20371. https://doi.org/10.48084/etasr.8994

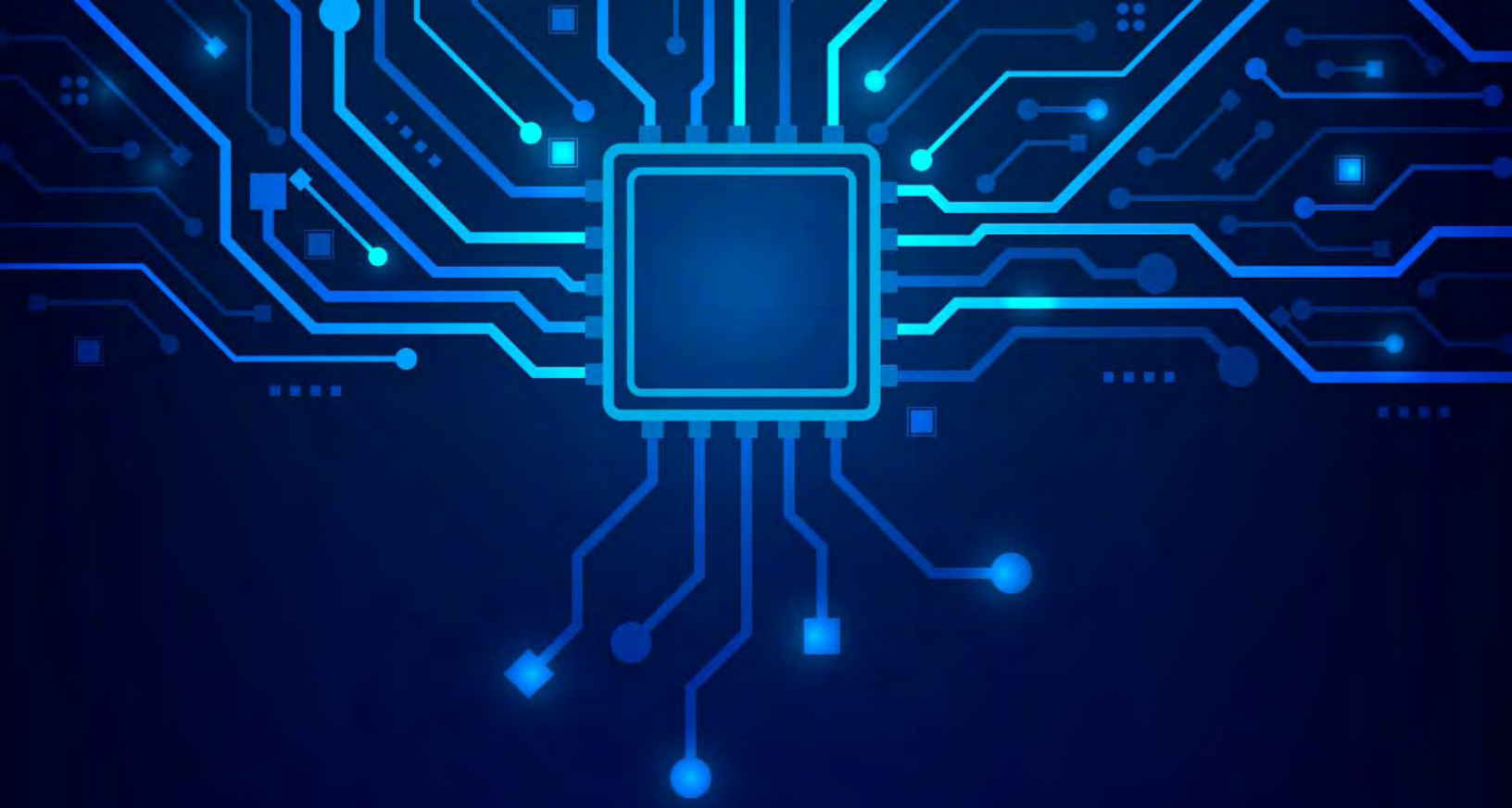
Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Truong, M. C., & Le, P. L. (2024). Evaluating the importance of digital tools or approaches to hospital performance using the analytic hierarchy process (AHP)-Delphi approach. *Journal of Hospital Management and Health Policy, 8*(0). https://doi.org/10.21037/jhmhp-24-60

Workum, J. D., Meyfroidt, G., Bakker, J., Jung, C., Tobin, J. M., Gommers, D., Elbers, P. W. G., van der Hoeven, J. G., & Van Genderen, M. E. (2026). AI in critical care: A roadmap to the future. *Journal of Critical Care*, 91, 155262. https://doi.org/10.1016/j.jcrc.2025.155262

Xia, J., Gao, J., & Tang, W. (2016). Nosocomial infection and its molecular mechanisms of antibiotic resistance. *Bioscience Trends*, *10*(1), 14–21. https://doi.org/10.5582/bst.2016.01020

Yepes-Barreto, I., Martínez, L., Girala, M., Sánchez-Santos, R., Graz, F., Restrepo, J. C., Beltrán, Ó., Peralta, M., Ridruejo, E., Alvares-da-Silva, M., Díaz, J., Coli, L., Daza, J., Sáenz, E., Espinosa, N., Itzel, T., Teufel, A., & Turnes, J. (2026). Barreras y oportunidades para la investigación en inteligencia artificial aplicada a la salud en América Latina: Una perspectiva basada en el análisis DOFA. *Hepatología*, *7*(1), 32–43. https://doi.org/10.59093/27112330.164



16.

Escucha humana e inteligencia artificial: límites y alcances en la investigación musical

José Luis Bautista López
Universidad Autónoma de Querétaro, México
ORCID: 0000-0002-9135-2778

Guillermo Vargas Rodríguez
Universidad Autónoma de Querétaro, México
ORCID: 0009-0001-5864-5376

Luis Javier Raúl Obregón Herrin
Universidad Autónoma de Querétaro, México
ORCID: 0009-0001-4961-1975

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.16

16.

Escucha humana e inteligencia artificial: límites y alcances en la investigación musical

Introducción

Cuando se trabaja con inteligencia artificial (IA) en música, la tentación más común es deslizarse de: *esta IA detecta patrones*, a: *esta IA entiende música*. Ese salto, tan humano y tan seductor, es el punto frágil de este capítulo. En los últimos años, la investigación musical se ha visto atravesada por herramientas provenientes de la *Music Information Retrieval* (MIR) y del aprendizaje automático: sistemas para describir audio, segmentar, etiquetar, comparar estilos, modelar expectativas y, más recientemente, generar música simbólica o audio con niveles llamativos de coherencia local. La MIR nació con el impulso de organizar y navegar grandes colecciones musicales digitales (Downie, 2003) y se expandió hacia una caja de herramientas cada vez más potente para el análisis musical asistido por datos (Müller, 2015). Esos avances han facilitado tareas antes lentas o directamente imposibles. El problema apareció cuando se confundió *descripción* con *comprensión*, *predicción* con *significado*, *regularidad estadística* con *experiencia musical*.

Aquí es donde entra el concepto de *audiation*, de Gordon (2012), en el que la escucha musical no solo es recordar un sonido, sino pensarlo musicalmente: anticiparlo, sostenerlo, transformarlo, sentir cómo pide una resolución o cómo se resiste a ella. No es una metáfora amable, es una experiencia humana y cotidiana para cualquiera que haya tarareado en silencio o ensayado mentalmente un pasaje antes de tocarlo. La neurociencia y la psicología musical han mostrado, además, que este *oír por dentro* activa regiones auditivas incluso sin sonido externo (Goh et al., 2023; Zatorre & Halpern, 2005), lo que subraya que la música ocurre también como actividad interna, encarnada, con espesor cognitivo.

Mientras tanto, los modelos computacionales han avanzado con rapidez en el terreno de la expectativa auditiva: Pearce y Wiggins (2012) proponen enfoques probabilísticos con ambición explicativa; y en el terreno generativo, existen ya taxonomías y marcos compara-

tivos sólidos para sistemas que producen música verosímil a gran escala (Briot et al., 2020; Herremans et al., 2017; Huang et al., 2018). Estos trabajos muestran algo con claridad: la IA puede capturar regularidades musicales y generar resultados que *suenan a* un estilo. Pero aquí surge la pregunta incómoda: ¿estamos frente a comprensión o solo frente a verosimilitud?

Este texto parte de una inquietud epistemológica concreta: en la investigación musical asistida por IA suelen mezclarse dos niveles distintos: el operativo —el sistema produce resultados útiles—, y el epistémico —estos resultados equivalen a conocimiento musical—. La confusión entre ambos es peligrosa, en parte, porque los modelos actuales son extraordinariamente persuasivos, pero esa persuasión no siempre va acompañada de transparencia. En aprendizaje automático existe, de hecho, una discusión profunda sobre interpretabilidad y confianza. Lipton (2017) advierte que la interpretabilidad suele usarse como un término ambiguo. Mientras que propuestas como *LIME* (Ribeiro et al., 2016) o *Model Cards* (Mitchell et al., 2019) buscan hacer explícitos los límites y condiciones de uso de los modelos. Aunque estas discusiones suelen aparecer en medicina o justicia, la música no es un terreno epistemológicamente inocente, cuando la IA se usa para analizar repertorios, evaluar desempeños u orientar decisiones pedagógicas. La pregunta ¿qué significa *comprender* música? deja de ser abstracta.

El objetivo de esta introducción es delimitar el terreno con honestidad: no se trata de negar el valor de la IA en música, pero tampoco de atribuirle una mente musical que no posee. La IA puede ser una herramienta eficaz para describir, predecir y generar, pero la *audiation* humana, con su mezcla de expectativa, memoria, afecto e intención, sigue siendo otra cosa: una cosa viva. Si somos cuidadosos, esa diferencia no será un obstáculo sino una brújula. En lo que sigue, se examinan las tensiones entre audición humana y análisis algorítmico como un problema de fronteras epistemológicas: ¿qué inferimos?, ¿con qué justificación y con qué lenguaje?, y ¿por qué los resultados computacionales no deben confundirse automáticamente con significado musical ni con comprensión?

Desarrollo

Audiation y comprensión musical humana

Cuando se habla de *audiation* no se refiere a una habilidad decorativa ni a una metáfora pedagógica conveniente; se habla de un proceso cognitivo activo, complejo y profundamente humano: la capacidad de pensar música en ausencia del sonido físico, anticiparla,

transformarla, sentir su tensión interna y reconocer cuándo algo *tiene sentido* musicalmente, incluso antes de que ocurra. Cualquiera que haya estudiado música —o simplemente haya vivido con ella— ha experimentado esto sin ponerle nombre: repetir mentalmente un pasaje, corregir una afinación imaginada, o sentir que una frase musical ya concluyó… o todavía no concluye. Edwin Gordon (2012) fue uno de los primeros en insistir, con una claridad poco común, en que la *audiation* no es memoria pasiva ni eco mental, sino cognición musical en acto, para él, auditar implica organizar internamente los sonidos según relaciones tonales, métricas y formales aprendidas culturalmente; no es solo recordar qué suena, sino comprender cómo funciona musicalmente aquello que suena, esta distinción es crucial, porque desplaza el foco desde el objeto sonoro hacia el sujeto que comprende.

Desde la psicología cognitiva y la neurociencia, esta idea ha recibido apoyo empírico sólido. Estudios clásicos sobre *Musical Imagery* (Goh et al., 2023; Zatorre & Halpern, 2005) muestran que imaginar música activa regiones auditivas del cerebro de forma sistemática, aunque con patrones distintos a la percepción directa; más aún, investigaciones posteriores han mostrado que la *audiation* involucra redes asociadas con la predicción, la memoria de trabajo y la evaluación emocional, lo cual refuerza la idea de que no se trata de un simple *replay* interno (Herholz et al., 2012). Es que, en la experiencia musical humana, anticipar musicalmente no es un subproducto: es el corazón mismo de la comprensión. Leonard Meyer (2009) ya había advertido que el significado musical emerge de la tensión entre expectativa y realización, entre lo que esperamos y lo que finalmente ocurre. Décadas después esta intuición fue formalizada desde enfoques contemporáneos que entienden la percepción musical como un proceso de inferencia predictiva, donde el oyente construye y actualiza modelos internos del flujo sonoro (Huron, 2006; Pearce, 2018). Aquí conviene subrayarlo: estos modelos no son explícitos ni conscientes, pero sí están cargados de historia, cuerpo y afecto.

En este punto, suele aparecer una confusión peligrosa, dado que muchos sistemas de IA también trabajan con predicción —por ejemplo, modelos que estiman la probabilidad del siguiente evento musical en una secuencia—: se asume que ambos procesos son comparables, pero esta analogía, aunque tentadora, es incompleta. En los modelos humanos, la predicción está integrada a una experiencia situada: escuchar música implica riesgo, sorpresa, aburrimiento, placer, frustración; implica, además, una orientación intencional hacia el sonido, como diría Dennett (1989): hay aquí una postura intencional que no se deja reducir a cálculo. Desde la filosofía de la mente, esta diferencia se ha formulado en términos

de intencionalidad y representación. La *audiation* no solo manipula símbolos o regularidades sino está dirigida a algo y tiene un carácter fenomenológico.

Searle (1980) lo expresó de manera provocadora al distinguir entre manipulación sintáctica y comprensión semántica: un sistema puede operar correctamente sobre símbolos sin *entender* aquello que esos símbolos representan, aunque su famoso argumento del *cuarto chino* ha sido ampliamente debatido, sigue siendo útil como advertencia epistemológica: producir salidas adecuadas no equivale automáticamente a comprender. Esto no significa —conviene decirlo sin dramatismos— que la *audiation* sea un proceso misterioso o inalcanzable para el análisis científico, al contrario: modelos como el de expectativa auditiva basada en teoría de la información propuesto por Pearce y Wiggins (2012), muestran que es posible formalizar ciertos aspectos de la experiencia musical sin vaciarla de contenido psicológico. Sin embargo, incluso estos modelos —cuidadosos en su anclaje cognitivo— reconocen que operan como aproximaciones a procesos humanos, no como sustitutos ontológicos de la experiencia.

La diferencia entre reconocimiento y comprensión es, aquí, decisiva. Un oyente puede reconocer un estilo, anticipar una cadencia o detectar una anomalía rítmica porque ha internalizado, a lo largo del tiempo, un entramado de regularidades cargadas de sentido cultural. Esta comprensión está atravesada por el cuerpo (*tempo* sentido, gesto imaginado), por la memoria afectiva y por contextos de uso. Como señala Leman (2007), la cognición musical está profundamente encarnada: no ocurre solo *en la cabeza*, sino en la interacción entre percepción, acción y entorno; por eso, cuando más adelante se comparen estos procesos con el análisis algorítmico, no se hará para descalificar a la IA, sino para delimitar cuidadosamente qué tipo de conocimiento está en juego. La *audiation* no es superior por razones románticas, sino distinta por razones estructurales; ignorar esa diferencia no nos vuelve más científicos, nos vuelve menos precisos.

El análisis algorítmico y sus supuestos epistemológicos

Cuando se habla de análisis algorítmico en música, se suele hacer con una mezcla de entusiasmo y confianza técnica. No es para menos, hoy existen sistemas capaces de segmentar audio, estimar tempo, identificar acordes, detectar patrones melódicos, modelar estilos y hasta generar continuaciones musicales que, en una primera escucha, resultan sorprendentemente verosímiles. Estas herramientas funcionan rápido, a escala y con consistencia. El problema empieza cuando dejamos de preguntar ¿qué tipo de conocimiento

producen exactamente esos resultados? En términos operativos, la mayoría de los sistemas actuales de análisis musical basados en IA —particularmente en el campo de la MIR— trabajan mediante la extracción de características, la modelación estadística y la optimización de predicciones. Un ejemplo clásico es la estimación automática de acordes (*chord recognition*), donde modelos entrenados sobre grandes corpus etiquetados, aprenden a asociar patrones espectrales con clases armónicas. Sistemas como *Chordino*, *Madmom* o implementaciones basadas en *Hidden Markov Models,* primero, y redes neuronales profundas, después, ilustran bien esta lógica: el modelo no *oye* acordes, calcula probabilidades condicionadas a partir de datos previos (Müller, 2015).

Algo similar ocurre con modelos generativos contemporáneos. Arquitecturas como *Music Transformer* o sistemas basados en *Recurrent Neural Networks*, procesan secuencias musicales como cadenas de eventos discretos y aprenden distribuciones de probabilidad sobre *qué suele venir después* (Huang et al., 2018). Desde un punto de vista técnico, el logro es notable: el sistema puede capturar dependencias de largo alcance, repetir motivos y sostener una coherencia local aceptable, pero conviene decirlo con cuidado: lo que el modelo aprende no es significado musical, sino regularidad estadística.

Aquí aparece el primer supuesto epistemológico fuerte del análisis algorítmico: la idea de que modelar patrones equivale a explicar el fenómeno. Este supuesto no siempre se explicita, pero suele operar en silencio. Si el sistema predice bien, se asume que *entendió* algo relevante. Sin embargo, la capacidad predictiva, por sí sola, no garantiza comprensión explicativa (Lipton, 2017). Como advierte Bishop (2006): un modelo puede ajustar con gran precisión una distribución de datos y aun así no ofrecer una explicación significativa del proceso que los genera, lo que obliga a distinguir cuidadosamente entre rendimiento predictivo y entendimiento del fenómeno.

De hecho, muchos modelos actuales son deliberadamente opacos, las redes neuronales profundas sacrifican interpretabilidad en favor de rendimiento, y aunque existen esfuerzos por hacerlas más explicables —como técnicas de explicación local tipo LIME o SHAP—, estas explicaciones suelen ser *post hoc* y no necesariamente reflejan los mecanismos reales del modelo (Lipton, 2017; Ribeiro et al., 2016). En música, esto significa que podemos obtener una segmentación *correcta* o una generación plausible sin poder articular qué relaciones musicales —en sentido humano— están siendo capturadas. Esto nos obliga a distinguir, desde una perspectiva epistemológica entre correlación, representación y comprensión. La mayoría de los sistemas de IA musical operan en el nivel de la correlación: detectan regu-

laridades estadísticas entre variables observables, algunos modelos incorporan formas de representación interna, pero estas representaciones no son intencionales ni semánticas en el sentido fuerte; no *se refieren* a la música como experiencia; solo codifican regularidades útiles para minimizar una función de pérdida, como advierte Bender et al. (2021): estos sistemas pueden producir salidas coherentes sin poseer ningún modelo del mundo, ni del significado de aquello que generan.

Aquí conviene detenerse un momento. Cuando un sistema generativo produce una progresión armónica correcta, o una melodía que suena a cierto estilo, es tentador atribuirle comprensión ya que la verosimilitud estilística es profundamente persuasiva para los humanos, pero esta persuasión no debe confundirse con agencia cognitiva; la intencionalidad no emerge automáticamente de la complejidad computacional (Floridi, 2011; Searle, 1980), el modelo no espera una cadencia, no se sorprende, no evalúa tensiones; simplemente ajusta probabilidades.

Otro supuesto frecuente es el de la neutralidad del dato. Los sistemas de análisis algorítmico se presentan a menudo como objetivos porque *solo trabajan con datos*. Sin embargo, esos datos están siempre curados, etiquetados y contextualizados por decisiones humanas: ¿qué repertorios se incluyen?, ¿qué estilos se consideran normativos?, ¿qué errores se toleran?, ¿qué métrica se optimiza? En música esto tiene consecuencias importantes ya que un modelo entrenado mayoritariamente con repertorio tonal occidental no solo aprende regularidades, aprende también una ontología implícita de lo que *cuenta* como música bien formada. Desde la epistemología de la IA, este fenómeno se ha discutido como *embedded values* o sesgos estructurales del diseño (Floridi et al., 2018).

Además, el análisis algorítmico tiende a descontextualizar, trabaja con fragmentos, ventanas temporales, vectores de características, eso es una fortaleza técnica, pero también una limitación conceptual, la música ocurre en situaciones: conciertos, ensayos, rituales, aulas, cuerpos en movimiento (Small, 1999). El algoritmo, en cambio, opera sobre representaciones despojadas de intencionalidad, corporeidad y contexto social, como señala Dreyfus (1992): los sistemas computacionales operan sobre representaciones formales en dominios cerrados y, por diseño, carecen del trasfondo situacional y del saber implícito que constituye el contexto del entendimiento humano. Pero nada de esto invalida el análisis algorítmico; la IA puede revelar regularidades que el oído humano no detecta fácilmente, comparar grandes *corpus* con una consistencia imposible para un investigador individual y abrir preguntas nuevas sobre estilo, estructura y variación, aunque su potencia heurística depende de que no se le exija aquello que no puede ofrecer, ya que cuando se confunde

correlación con *sentido*, o *predicción* con *comprensión*, no se amplía el conocimiento musical, sino se empobrece conceptualmente. Por eso, el desafío no es elegir entre *audiation* o análisis algorítmico, sino articular sus diferencias sin jerarquizarlas falsamente, reconocer que el análisis algorítmico produce un tipo específico de conocimiento —formal, estadístico, operacional—: es el primer paso para integrarlo de manera epistemológicamente responsable en la investigación musical, solo así podremos usar estas herramientas sin convertirlas, por comodidad o fascinación, en sujetos cognitivos que nunca han sido.

Puntos de fricción entre audiation y análisis algorítmico

Llegados a este punto, la fricción ya no es técnica; es conceptual y, en cierto modo, también ética. La *audiation* y el análisis algorítmico no chocan porque uno sea mejor que el otro, sino porque operan con supuestos distintos sobre qué cuenta como música y como conocimiento musical; cuando estos supuestos se confunden, el diálogo se vuelve ruido.

El primer punto de fricción aparece en la reducción de la experiencia musical a datos. Para que un sistema algorítmico funcione, la música debe volverse representable: números, vectores, etiquetas, secuencias discretas. Este paso no es trivial ni inocente: implica decidir qué aspectos del fenómeno son relevantes y cuáles pueden descartarse sin *perder demasiado*. En el análisis de audio, por ejemplo, se privilegian descriptores espectrales, temporales o armónicos; en la música simbólica, se abstraen eventos como alturas, duraciones y dinámicas, todo lo demás —gesto, intención, contexto performativo, afecto— queda fuera o, en el mejor de los casos, implícito. Desde la *audiation*, en cambio, esos elementos no son periféricos, son constitutivos. Cuando un músico *oye* internamente una frase, no imagina solo alturas y duraciones: imagina peso, dirección, respiración, incluso el esfuerzo corporal necesario para producir ese sonido. Tal como ha mostrado la literatura sobre cognición musical encarnada, la comprensión musical está íntimamente ligada a esquemas motores y sensoriomotrices (Leman, 2007); reducir la música a datos puede ser metodológicamente útil, pero confundir esa reducción con la experiencia misma es un error categorial.

Un segundo punto de fricción surge en la confusión entre descripción formal y sentido musical. Muchos sistemas algorítmicos describen con gran precisión qué ocurre en una pieza: ¿cuántas veces aparece un motivo?, ¿qué progresiones armónicas son más frecuentes?, ¿qué patrones rítmicos dominan un estilo?, pero el sentido musical —lo que *significa* una repetición, ¿por qué una cadencia conmueve o decepciona? — no se agota en esa descripción. En la *audiation* el sentido emerge de la relación entre expectativa, me-

moria y afecto, no basta con que algo ocurra muchas veces; importa cómo ocurre y qué hace sentir cuando ocurre. Aquí conviene recordar una advertencia clásica: explicar no es lo mismo que describir; incluso afirmaciones verdaderas y precisas pueden quedarse en *mera descripción* y no constituir una explicación del fenómeno (Woodward & Ross, 2003). En música, esta distinción es especialmente delicada porque la descripción formal puede parecer, superficialmente, una explicación suficiente, pero cuando un sistema detecta que cierto acorde suele seguir a otro, no está explicando por qué ese movimiento genera cierre o tensión; solo está constatando una coocurrencia.

El tercer punto de fricción tiene que ver con la ausencia de experiencia e intencionalidad en los sistemas algorítmicos. Ya se ha señalado antes, pero aquí se vuelve central: la *audiation* implica *estar dirigido a*. El oyente espera, se equivoca, corrige, se sorprende; el sistema algorítmico, en cambio, no espera nada, no tiene expectativas en el sentido fenomenológico: solo calcula probabilidades. Incluso cuando se habla de *modelos de expectativa*, conviene no olvidar que se trata de metáforas operativas, no de estados mentales. Desde la epistemología de la IA, este problema se ha formulado como una tensión entre simulación y realización; un sistema puede simular comportamientos asociados a la comprensión sin realizar realmente los procesos que, en humanos, constituyen dicha comprensión (Floridi, 2011). En música, esta simulación puede ser tan convincente que incluso músicos experimentados atribuyen intención donde solo hay optimización estadística.

Un cuarto punto de fricción —menos discutido, pero igualmente importante— aparece en los riesgos metodológicos de la investigación musical asistida por IA: cuando el análisis algorítmico se toma como árbitro final, existe la tentación de ajustar las preguntas de investigación a aquello que el sistema puede responder. Poco a poco, sin mala intención, el objeto de estudio se estrecha, se investiga lo que es computable, no necesariamente lo que es musicalmente relevante. Este fenómeno, descrito en otros campos como *methodological lock-in*, puede empobrecer el horizonte epistemológico de la disciplina (Selbst et al., 2019). Además, los resultados algorítmicos suelen adquirir una autoridad indebida por su apariencia de objetividad; un gráfico, una métrica de precisión, una matriz de confusión pueden imponerse fácilmente sobre descripciones cualitativas o análisis fenomenológicos. No todo lo que importa se deja contar, y no todo lo que se puede contar importa por igual.

Finalmente, hay una fricción más sutil: la temporalidad vivida frente a la temporalidad computada. Mientras que la *audiation* ocurre en un tiempo subjetivo, elástico, cargado de anticipación y memoria, el análisis algorítmico, en cambio, segmenta el tiempo en ventanas,

frames, eventos discretos; esta segmentación es necesaria para el cálculo, pero transforma radicalmente el fenómeno, como ha señalado Varela (1999): el tiempo vivido no es reducible a una sucesión de instantes medibles sin perder algo esencial.

Reconocer estos puntos de fricción no implica renunciar al uso de la IA en investigación musical; implica, más bien, usar estas herramientas con conciencia de sus límites, evitando, tanto la fascinación acrítica, como el rechazo defensivo. La *audiation* y el análisis algorítmico pueden dialogar, pero solo si aceptamos que hablan lenguajes distintos y que ninguno agota, por sí solo, el fenómeno musical; en esa tensión —incómoda, productiva— se juega buena parte del futuro de la investigación musical contemporánea.

La inteligencia artificial como herramienta, no como sujeto cognitivo

Llegados aquí, la pregunta ya no es si la IA sirve para la investigación musical —eso está fuera de duda—, sino cómo debe situarse epistemológicamente para no distorsionar aquello que pretende iluminar. Y la respuesta que se propone, sin rodeos, es esta: la IA funciona de manera más fértil cuando se la entiende como herramienta cognitiva ampliada, no como sujeto cognitivo sustituto. Parece una distinción obvia, pero la verdad es que no siempre lo es en la práctica. En el trabajo cotidiano con sistemas algorítmicos resulta fácil deslizarse hacia un lenguaje que atribuye agencia: *el modelo descubre*, *la red entiende*, *el sistema decide*.

Estas expresiones no son solo metáforas inocentes; configuran expectativas epistemológicas. Desde la filosofía de la tecnología se ha advertido que este tipo de antropomorfización no describe lo que los sistemas son, sino lo que los humanos proyectamos sobre ellos (Floridi, 2014). En música, esta proyección es especialmente intensa porque el resultado puede sonar expresivo, coherente, incluso *emocional*, pero que algo suene expresivo no significa que lo sea desde el punto de vista cognitivo. Entender la IA como herramienta implica reconocer que su potencia reside en su capacidad para extender ciertas operaciones humanas, no para reemplazar la experiencia musical.

Un sistema puede analizar miles de piezas, detectar regularidades invisibles al oído individual, proponer hipótesis formales o generar material que funcione como estímulo creativo. En ese sentido, la IA actúa como un amplificador heurístico, algo similar a lo que ocurrió cuando las notaciones musicales se complejizaron, o cuando el análisis schenkeriano ofreció nuevas formas de ver la estructura tonal: no sustituyeron la escucha, la reorientaron. Este enfoque instrumental se alinea con una concepción pragmática del conocimiento científico.

Andy Clark (2011) sugiere, desde la teoría de la mente extendida, que las herramientas no están *fuera* del proceso cognitivo, forman parte de él siempre que haya un sujeto que interprete, evalúe y actúe sobre sus resultados. La diferencia crucial es que, en este marco, la agencia sigue siendo humana, el modelo no formula preguntas, no revisa sus supuestos, no duda de sus resultados, esa asimetría no es un defecto, es una condición estructural; además, concebir la IA como herramienta obliga a una responsabilidad epistemológica explícita, si el sistema no es un sujeto, entonces alguien debe responder por la selección de los datos, la formulación de las tareas, la interpretación de las salidas y las inferencias que se derivan de ellas. Esta idea conecta con propuestas recientes en ética y gobernanza de la IA, que insisten en mantener una línea clara de responsabilidad humana en el uso de sistemas algorítmicos (Floridi et al., 2018; Mitchell et al., 2019). En investigación musical, esto significa no esconder decisiones interpretativas detrás de métricas técnicas ni presentar resultados como descubrimientos autónomos del sistema.

Hay, además, un beneficio pedagógico y metodológico en esta postura: cuando la IA se presenta como herramienta, se vuelve discutible, criticable, incluso falible; se puede preguntar: ¿qué deja fuera este modelo? ¿qué asume sobre la música? ¿qué tipo de escucha privilegia? ¿qué experiencias musicales invisibiliza? Estas preguntas no debilitan la investigación, la fortalecen. En cambio, cuando la IA se presenta como sujeto —como si *supiera* música—, el espacio crítico se reduce, y con él, la posibilidad de aprendizaje. Desde la música, esta distinción tiene consecuencias claras, la *audiation* no compite con la IA, dialoga con ella desde otro plano; el análisis algorítmico puede sugerir regularidades, pero es la *audiation* la que evalúa si esas regularidades tienen sentido musical, si resuenan con una práctica, un estilo, una experiencia encarnada. En ese diálogo, la IA no dicta, propone; no comprende, asiste. Y quizá ahí reside su lugar más honesto y más productivo.

Defender que la IA no es un sujeto cognitivo no es restarle valor, sino ubicarla con precisión; como herramienta, puede enriquecer la investigación musical, ampliar el campo de lo observable y tensionar nuestras categorías analíticas; como supuesto sujeto, en cambio, corre el riesgo de oscurecer aquello que hace de la música una experiencia humana irreductible. Mantener esta distinción viva no es un gesto conservador; es, paradójicamente, una condición para innovar sin perder el norte.

Conclusiones

A lo largo de este ensayo se ha intentado sostener una idea sencilla, aunque incómoda: no todo lo que la IA hace bien equivale a comprensión musical, y no toda comprensión musical es reducible a lo que la IA puede formalizar. La fricción entre *audiation* y análisis algorítmico no es un problema a resolver eliminando un polo; es una tensión a habitar con cuidado. La *audiation* —escucha interna, anticipación, memoria afectiva y orientación intencional— sigue siendo el núcleo de la comprensión musical humana, no porque sea misteriosa o romántica, sino porque integra dimensiones que no se dejan traducir sin pérdida: cuerpo, experiencia situada, expectativa vivida y sentido cultural. El análisis algorítmico, por su parte, produce un conocimiento distinto: formal, estadístico, operacional, potente para describir regularidades, comparar grandes *corpus* y generar hipótesis. Confundir estos niveles —tratar la correlación como significado o la predicción como comprensión— no amplía el conocimiento musical, lo empobrece; de ahí la propuesta central: ubicar a la IA como herramienta y mediador cognitivo, no como sujeto. Esta delimitación no resta valor a la tecnología; al contrario, la vuelve más útil.

Como herramienta, la IA amplifica capacidades humanas, acelera procesos, revela patrones y tensiona nuestras categorías analíticas; como supuesto sujeto cognitivo, en cambio, introduce ambigüedad epistemológica, diluye responsabilidades interpretativas y favorece una antropomorfización que oscurece más de lo que aclara. Esta distinción tiene implicaciones prácticas claras para la investigación musical. En el plano metodológico, exige explicitar supuestos: qué se modela, qué queda fuera, qué tipo de conocimiento se reclama y con qué justificación; en el plano interpretativo, obliga a mantener la agencia humana en la formulación de preguntas, la lectura de resultados y la evaluación de su relevancia musical; y en el plano educativo, invita a presentar la IA no como reemplazo de la escucha, sino como andamiaje para escuchar mejor, comparar con mayor perspectiva y pensar con más herramientas.

Quizá el gesto más importante sea aceptar que la innovación no consiste en delegar la comprensión, sino en redistribuir inteligentemente las tareas, dejar a la IA aquello que hace bien —procesar grandes volúmenes, detectar regularidades, generar variantes— y preservar para la *audiation* aquello que le es propio —sentido, expectativa vivida, juicio musical— no es una concesión conservadora, es una estrategia epistemológica responsable. En última instancia, este trabajo no propone una frontera rígida, sino una cartografía de límites: una guía para dialogar sin confundir lenguajes, colaborar sin atribuir mentes donde no las hay,

y avanzar sin perder de vista que la música, incluso en tiempos de algoritmos sofisticados, sigue siendo una experiencia profundamente humana, mantener esa claridad no frena la investigación musical contemporánea: le da suelo, y, a veces, eso es justo lo que hace falta para seguir caminando.

Referencias

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21*, 610–623. https://doi.org/10.1145/3442188.3445922

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Briot, J.-P., Hadjeres, G., & Pachet, F. (2020). *Deep Learning Techniques for Music Generation*. Springer Nature Switzerland.

Clark, A. (2011). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford Univ. Press.

Dennett, D. C. (1989). *The intentional stance*. MIT Press. http://archive.org/details/intentionalstanc00dani

Downie, J. S. (2003). Music information retrieval. A*nnual Review of Information Science and Technology*, *37*(1), 295–340. https://doi.org/10.1002/aris.1440370108

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.

Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199232383.001.0001

Floridi, L. (2014). *The Fourth Revolution: How the infosphere is reshaping human reality.* Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Goh, R. Z., Phillips, I. B., & Firestone, C. (2023). The perception of silence. *Proceedings of the National Academy of Sciences*, *120*(29), e2301463120. https://doi.org/10.1073/pnas.2301463120

Gordon, E. (2012). *Learning sequences in music: Skill, content, and patterns: A music learning theory*. GIA Publication.

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal Correlates of Perception, Imagery, and Memory for Familiar Tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382–1397. https://doi.org/10.1162/jocn_a_00216

Herremans, D., Chuan, C.-H., & Chew, E. (2017). A Functional Taxonomy of Music Generation Systems. *ACM Comput. Surv.*, *50*(5), 69:1-69:30. https://doi.org/10.1145/3108242

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman, M. D., Dinculescu, M., & Eck, D. (2018). *Music Transformer* (arXiv:1809.04281). arXiv. https://doi.org/10.48550/arXiv.1809.04281

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. The MIT Press. https://doi.org/10.7551/mitpress/7476.001.0001

Lipton, Z. C. (2017). The Mythos of Model Interpretability (arXiv:1606.03490). *arXiv*. https://doi.org/10.48550/arXiv.1606.03490

Meyer, L. (2009). *La emoción y el significado en la música*. Alianza Musical.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19*, 220–229. https://doi.org/10.1145/3287560.3287596

Müller, M. (2015). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer International Publishing. https://doi.org/10.1007/978-3-319-21945-5

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

Pearce, M. T., & Wiggins, G. A. (2012). Auditory Expectation: The Information Dynamics of Music Perception and Cognition. *Topics in Cognitive Science*, *4*(4), 625–652. https://doi.org/10.1111/j.1756-8765.2012.01214.x

Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. En J. DeNero, M. Finlayson, & S. Reddy (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 97–101). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-3020

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

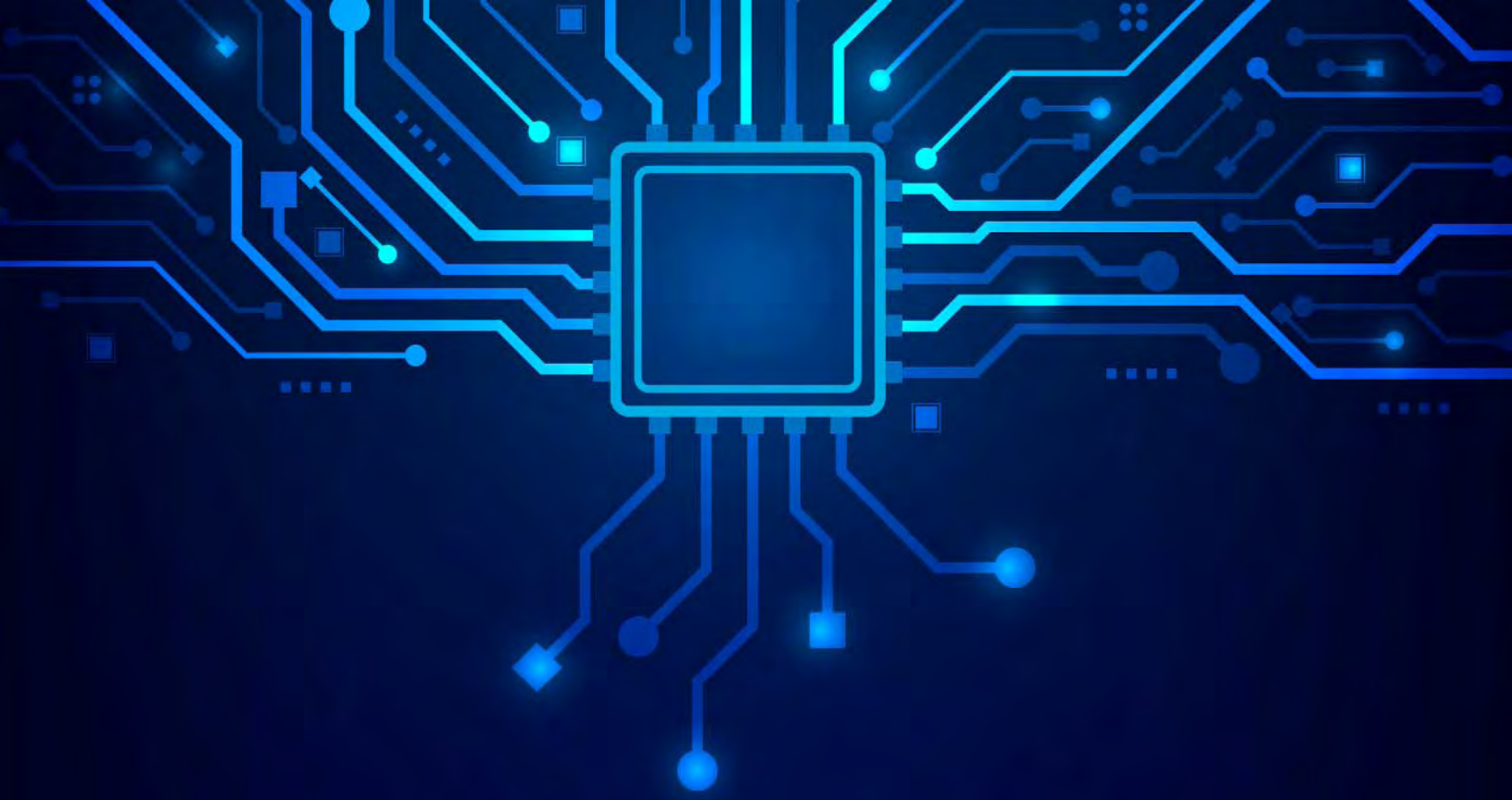
Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19*, 59–68. https://doi.org/10.1145/3287560.3287598

Small, C. (1999). El Musicar: Un ritual en el Espacio Social. *TRANS–Revista Transcultural de Música*, 4. https://www.sibetrans.com/trans/articulo/252/el-musicar-un-ritual-en-el-espacio-social

Varela, F. J. (1999). *Ethical know-how: Action, wisdom, and cognition* (pp. ix, 85). Stanford University Press.

Woodward, J., & Ross, L. (2003). *Scientific Explanatio*n. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/fall2023/entries/scientific-explanation/

Zatorre, R. J., & Halpern, A. R. (2005). Mental Concerts: Musical Imagery and Auditory Cortex. *Neuron*, *47*(1), 9–12. https://doi.org/10.1016/j.neuron.2005.06.013



17.

Evolución de la representación gráfica: del trazo humano al algoritmo

Luz Angélica Mondragón del Angel

Universidad Autónoma de Querétaro, México

ORCID: 0000-0001-5997-182X

Inés Guadalupe Germán Aguilar

Universidad Autónoma de Querétaro, México

ORCID 0009-0003-1707-0734

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.17

17.

Evolución de la representación gráfica: del trazo humano al algoritmo

Introducción

La representación gráfica ha sido, desde los orígenes de la humanidad, una herramienta de comunicación muy importante para la comprensión y transmisión de formas de vida y conocimiento. El manejo de imágenes, símbolos y estructuras diagramadas ha dado a conocer e interpretar la realidad. Asimismo, ha permitido organizar la información y las ideas de manera visual antes del desarrollo del lenguaje escrito. De esta manera, las civilizaciones a través del tiempo han tenido una evolución que ha dado lugar a procesos y sistemas estructurados, logrando adquirir un papel esencial en el desarrollo científico de disciplinas como la ingeniería, la arquitectura y el diseño. Estos mecanismos de comunicación gráfica se convierten en un lenguaje universal capaz de facilitar la comprensión y la resolución de problemas.

En este sentido, la evolución de la representación gráfica en el tiempo, su sistematización, la utilización de herramientas computacionales para la representación digital y, ahora, la irrupción de la inteligencia artificial (IA), han permitido el manejo de la información a través de gráficos dinámicos y modelos tridimensionales que mejoran la interpretación visual. Estamos ante una transformación en el ámbito de la comunicación con el uso de la IA, al integrar capacidades de representación, producción y adaptación automática. Sin embargo, estas características han generado tensiones en relación con su uso indiscriminado que puede derivar en la producción de representaciones visuales carentes de rigor técnico y significado contextual, como señala Martínez (2025, p. 87) respecto a la cada vez más extendida utilización de imágenes arquitectónicas obtenidas por medio de la IA “estas herramientas de IA han segregado la resolución de problemas técnico-constructivos al decantarse por exploraciones ilustrativas, es decir, la arquitectura se ha supeditado a gráficos”.

Diversos autores coinciden en que la integración de la IA en el ámbito de la arquitectura marca cambios importantes en los procesos de diseño que reinventarán la profesión (Cutieru, 2021; Martínez 2025). El paso de las técnicas manuales al diseño asistido por computadora (CAD, por las siglas en inglés de *Computer Aided Design*) y posteriormente, a

los sistemas paramétricos y generativos, ha convertido la representación en un instrumento operativo y dinámico capaz de optimizar tareas complejas y guiar la toma de decisiones de diseño, mismas que conllevan a nuevas tendencias, a una nueva forma de pensar y hacer arquitectura. Por lo tanto, el objetivo de este ensayo es analizar la evolución histórica y técnica de la representación gráfica, examinando sus distintos usos y enfoques, con el fin de comprender su estado actual en relación con tecnologías digitales como la IA, así como los desafíos y oportunidades que presenta esta herramienta como medio fundamental para el desarrollo de proyectos en la disciplina arquitectónica.

Desarrollo

El hombre, las primeras civilizaciones y la necesidad de comunicar

La representación gráfica nace de la necesidad de comunicación visual antes del lenguaje escrito. En la era prehistórica se conocen pinturas rupestres (Figura 1) que datan de 4,000 A.C. en Altamira, España y en Lascaux, Francia, las cuales describen a manera de narración gráfica rituales de cómo el ser humano cazaba animales para sobrevivir. Los primeros registros gráficos también se relacionan con el culto de los antepasados y la fertilidad. Estas representaciones de símbolos, figuras humanas y animales son el primer intento de abstracción visual de la realidad, ya que representan información de la forma y estilo de vida a través del ser humano por medio de imágenes. En consecuencia, la evolución histórica del dibujo se basa en el desarrollo de sistemas gráficos orientados a transmitir un amplio rango de ideas, por medio de la conjugación de signos simbólicos y representativos. Este proceso condujo a incluir figuras con valor fonético en los jeroglíficos, marcando así una transición de la escritura ideográfica a la fonográfica (Gelb, 1987).

Figura 1

Pintura rupestre de la cueva de Altamira, España

Nota. Tomado de Rameessos (2015).

La escritura jeroglífica (Figura 2) se estableció como un sistema de comunicación efectivo a medida que las distintas civilizaciones fueron desarrollándose, su consolidación permitió implementarse en los ámbitos pedagógico, administrativo y religioso, diversificando sus funciones dentro de la estructura social. Ejemplo de ello son Egipto, Mesopotamia y Grecia. En Egipto se tienen evidencias de jeroglíficos con proporciones geométricas estrictas y planos de espacios religiosos y funerarios (templos y pirámides). En Mesopotamia se evidencia una escritura cuneiforme, tablas de arcilla con registros económicos y diagramas astronómicos. Y Grecia por su parte muestra estudios de geometría con Euclides, representación matemática científica con Pitágoras y los inicios de una metodología del dibujo como herramienta de razonamiento. Sin embargo, la representación de narrativas mitológicas o políticas rara vez atendía a la fidelidad visual, más bien hacían referencia a un canon esquemático, bajo el cual la disposición y la dimensión de figuras ocurría según su importancia relativa en el contexto de la narrativa o del tema representado, sin considerar la parte tridimensional de los mismos (Villanueva, 2005).

Figura 2

Alfabeto Tell Umm-el Marra, Siria. 2,400 AC

Nota. Tomado de Schwartz (2021).

Transmisión del conocimiento en la Edad Media y la revolución visual y científica del Renacimiento

En el periodo Antiguo, las representaciones gráficas producidas son imprecisas, sin una organización definida, porque todavía no se establece el concepto de campo pictórico en ellas, lo que dificulta ubicar las figuras con una lógica espacial coherente. Durante este extenso período histórico que precedió al Renacimiento italiano, se pueden identificar tres fases importantes en la representación gráfica: el modo Egipto-mesopotámico, el grecorromano y el medieval (Hauser, 1972; Gombrich, 1986).

Durante la Edad Media se buscó reflejar el conocimiento a través de la representación visual enfocada en valores espirituales y doctrinales (Panofsky, 1991), no como un interés por la exactitud de la representación, sino por mostrar y enseñar el conocimiento, ejemplo de ello es la representación arquitectónica mediante dibujos y trazos para la edificación de catedrales góticas (Figura 3) que respondían a sistemas geométricos prácticos basados en proporciones simples y figuras reguladoras. Se tienen evidencias de ilustraciones en manuscritos y mapas medievales de territorios dominados, diagramas religiosos y filosóficos. En la época medieval el objetivo principal de las imágenes no era reproducir la realidad de manera exacta, sino transmitir significados priorizando la representación basada en modelos o estándares preestablecidos (Panofsky, 1991). Al final del Medioevo, un desarrollo evidente de las técnicas constructivas no se refleja aún en la representación gráfica, esto se justifica porque los profesionales responsables de las obras arquitectónicas eran normalmente los mismos que las concebían.

Figura 3

Fachada gótica dibujada por Villard de Honnecourt (siglo XIII)

Nota. Tomada de Honnecourt (ca. 1230).

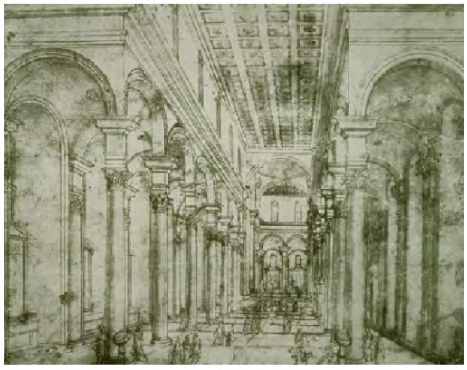
En el periodo del Renacimiento se consolida la relación entre arte, ciencia y técnica, ya que se empieza a desarrollar una división franca entre las necesidades estéticas, las de conocimiento científico y las de los procesos técnicos, buscando la precisión en los trabajos desarrollados en cada rubro. Se desarrolla la perspectiva científica de Brunelleschi, los estudios de anatomías de Leonardo da Vinci y los procesos de dibujo técnico con la representación del espacio y la forma de manera precisa. Este avance supuso un cambio epistemológico en la manera de concebir la imagen y su relación con la realidad. Panofsky (1991) sostiene que durante la Edad Media no fue posible la consolidación de un sistema

de representación basado en la concepción de una distancia objetiva entre el observador y el objeto representado, condición fundamental para el surgimiento de la perspectiva lineal.

En los siglos XIV y XV, en Europa Occidental, el interés por conseguir una representación visual más exacta, la observación de la vida diaria y del medio natural, así como los estudios científicos, propiciaron una nueva modalidad de representación conocida como *renacentista* (Panosfky,1991). Con este enfoque, el plano pictórico se organizó con límites más precisos y claros, y la imagen empezó a ser vista como una ventana a través de la cual se transmite la información, incorporando un sentido de profundidad basado en una visión monocular similar a la humana. Fue entonces a principios del siglo XV, cuando se fijaron los fundamentos de la perspectiva cónica (Figura 4). Gombrich (1995) atribuye a Brunelleschi el primer experimento sistemático de demostración de la perspectiva cónica.

Figura 4

Interior de San Paolo fuori le Mura en Roma



Nota. Tomada de Palladio (1570).

La Revolución Industrial, la Geometría Descriptiva y la evolución al siglo XX

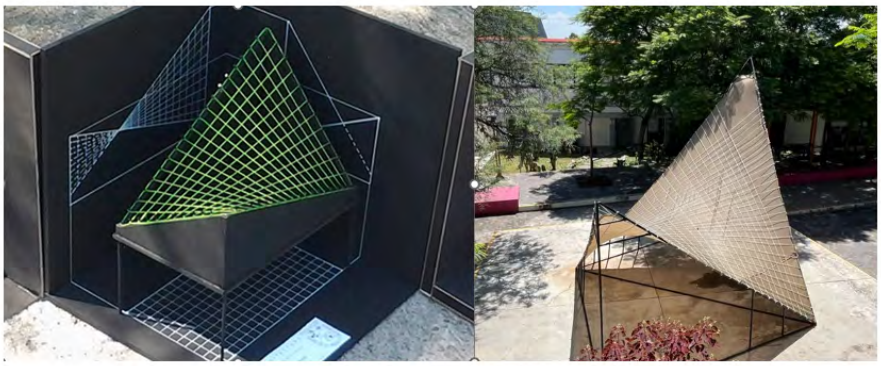
A fines del siglo XVIII, los sistemas de representación gráfica, así como los métodos geométricos y matemáticos para describir los sólidos, se encontraban ya plenamente sistematizados. Con la invención de la máquina de vapor y los procesos de producción en serie, la representación gráfica se convirtió en un recurso muy importante para la industria y la ingeniería. Ante la necesidad de replicar los diseños y procesos constructivos de carácter militar principalmente, Gaspar Monge comienza el desarrollo de una metodología de dibujo que pudiera ser comprendida en diferentes disciplinas y replicada de manera sencilla en la resolución de problemas (Villanueva, 2005). Monge es reconocido como el principal precursor

de la geometría descriptiva al establecerla, a finales del siglo XVIII, como un método sistemático para representar objetos tridimensionales mediante proyecciones bidimensionales.

La aportación de Monge consistió en la formalización de procedimientos gráficos que permitieran definir con precisión la forma y posición de los cuerpos en el espacio, sentando las bases para la comunicación técnica en la ingeniería y la arquitectura. A través de este enfoque, la representación dejó de depender de recursos empíricos o artísticos para convertirse en un lenguaje riguroso, capaz de describir, analizar y resolver problemas espaciales complejos, lo que consolidó a la geometría descriptiva como una disciplina esencial en la formación técnica moderna (Monge, 1799). Esta metodología geométrica descriptiva genera un cambio importante en la manera de representar los elementos que deben ser construidos y replicados por personas ajenas a la creación del proyecto. Esto se debe al claro manejo de un lenguaje universal matemático de representación que permite su interpretación para la resolución de problemas entre cuerpo y espacio, por medio de la descripción de objetos que pueden ser identificados con claridad en un espacio limitado por tres planos de representación para ser construidos con precisión (Figura 5).

Figura 5

Geometría descriptiva aplicada a la arquitectura



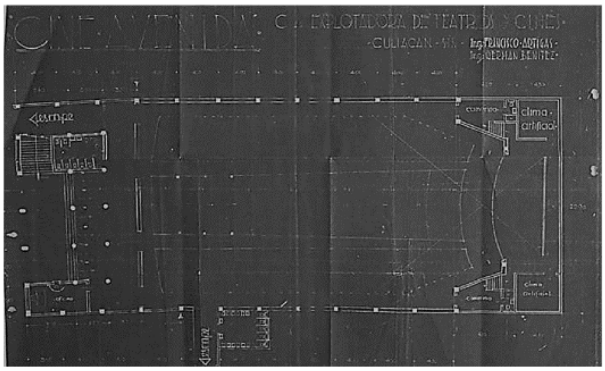
El uso de proyecciones ortogonales, del sistema de planos acotados y de la perspectiva axonométrica, cuyo desarrollo se había iniciado desde el siglo XVI, permitió la consolidación del dibujo técnico como un instrumento normativo y de organización, estrechamente vinculado a las nuevas demandas industriales. La difusión de numerosos tratados especializados, junto con la institucionalización de la enseñanza de las artes y los oficios en academias, contribuyó decisivamente a la estructuración y transmisión de este cuerpo de conocimientos. Asimismo, la expansión de estos sistemas, textos e instituciones en Europa y América

evidenció la incorporación gradual de nuevos valores económicos, sociales, filosóficos y estéticos en los procesos de representación, estableciendo una correspondencia directa entre estos y la sociedad surgida con la Revolución Industrial (Villanueva, 2005).

A partir de este hecho, se inicia la realización de planos mecánicos, constructivos y arquitectónicos de espacios de manera técnica estandarizada. Al trabajar bajo una metodología se normalizan símbolos, escalas, reglas de representación en el diseño, fabricación y construcción de objetos y espacios arquitectónicos. En este sentido, el desarrollo de la máquina de reproducción heliográfica marcó un punto de partida en la representación arquitectónica al transformar radicalmente la manera en que los planos eran producidos y distribuidos. El perfeccionamiento de la cianotipia por John Herschel en 1842 permitió obtener copias fieles y económicas de dibujos técnicos (Figura 6) mediante un proceso fotosensible que generaba imágenes en fondo azul con líneas blancas (Ware, 1999).

Figura 6

Copia heliográfica del plano original de proyecto de cine, diseño de los arquitectos Artigas y Benítez



Nota. Tomado de Mendoza et al. (2025).

Este procedimiento sustituyó la copia manual, posibilitando la reproducción masiva de planos idénticos y favoreciendo la estandarización de la información gráfica entre arquitectos, ingenieros y constructores. Como consecuencia, el dibujo con lenguaje geométrico descriptivo dejó de ser únicamente un instrumento creativo para convertirse en un documento técnico, contractual y operativo dentro de los procesos industriales. Además, la facilidad de reproducción contribuyó a la rápida circulación internacional de proyectos ferroviarios, urbanos e industriales, consolidando el dibujo técnico como un lenguaje universal acorde con las dinámicas productivas y organizativas de la modernidad industrial (Carpo, 2001).

Era digital contemporánea

El tránsito de la máquina copiadora heliográfica a la computadora representa una de las transformaciones más importantes de la historia de la representación gráfica arquitectónica. Mientras que la heliografía permitió la reproducción masiva y estandarizada de planos durante la Revolución Industrial, consolidando el dibujo técnico como documento contractual y medio de coordinación industrial, la incorporación de la computadora en el siglo XX redefinió el propio proceso de diseño. Con el desarrollo de los primeros sistemas CAD en las décadas de 1960 y 1970, el dibujo dejó de ser un registro estático para convertirse en un entorno dinámico de modelado y modificación continua (Negroponte, 1970). La representación digital no solo facilitó la precisión geométrica y la edición inmediata, sino que introdujo nuevas posibilidades como la modelación tridimensional, la simulación estructural y la visualización en tiempo real.

Entre los años 1950 y 1970, antes de la digitalización masiva, la representación gráfica se basaba en técnicas manuales de dibujo y composición. Posteriormente, la representación gráfica evolucionó de ser un proceso artesanal y físico a un ecosistema digital interactivo, encaminado a los primeros pasos digitales. En 1960, Iván Sutherland creó *Sketchpad*, el primer programa que permitió la manipulación directa de objetos gráficos en pantalla, marcando el inicio del CAD. Con la llegada de la computadora personal cambió radicalmente la producción visual y la revolución del software. Con la introducción de *Apple Macintosh* y programas como *Adobe Illustrator* en 1987 y *Photoshop* en 1990, en los cuales se realiza Modelado 3D, infografías interactivas y la visualización de *big data* (Kolarevic, 2003).

Esto permitió a cualquier persona experimentar con herramientas que antes requerían talleres especializados. Se había llegado a la convergencia entre los aspectos artísticos y técnicos del dibujo, que hasta la actualidad ha estado permeado por el propio desarrollo tecnológico, en particular de las Tecnologías de la Información y Comunicación (TIC). Con la utilización de tecnologías digitales a partir de diversos medios para el dibujo técnico se logran obtener recreaciones virtuales en 3D (Figura 7), que muestran una representación de los objetos con características específicas para su construcción (por ejemplo, su verdadera magnitud y forma) que favorecen la percepción del espectador.

Figura 7

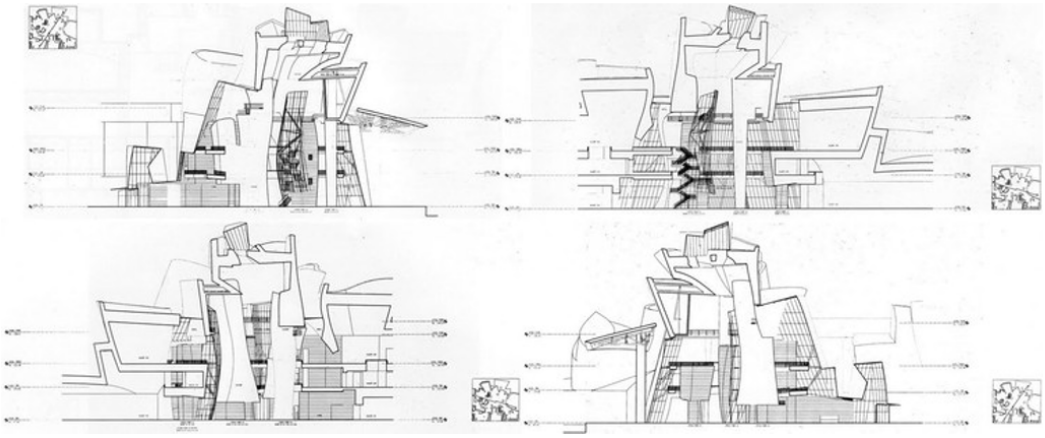
Modelo digital 3D Sketchup

Nota. Tomado de Sketchup (2026).

Hacia los años 90, con el auge de Internet, la representación gráfica dejó de ser puramente impresa. Surgió la necesidad de diseñar interfaces de usuario y experiencias interactivas que funcionaran en pantallas. En 1997 se desarrolló la primera obra arquitectónica a partir de tecnología digital: el museo Guggenheim de Bilbao (Figura 8) a cargo del arquitecto Frank Ghery, quien utilizó una tecnología de escaneo 3D basada en ingeniería inversa para capturar las características superficiales de una maqueta física y convertirlas en coordenadas tridimensionales, llamada *nube de puntos*. Mediante una interfaz gráfica, los puntos se transformaron en líneas, superficies y volúmenes, corrigiendo errores hasta aproximarse a la forma de la maqueta. Posteriormente, la información se trasladó al *software* de modelado *CATIA* para procesarla, ajustar irregularidades y construir la geometría del edificio, obteniendo finalmente un modelo digital preciso (Fraile, 2018).

Figura 8

Alzados y secciones del Museo Guggenheim de Bilbao



Nota. Tomado de Pagnotta (2015).

Durante los años 2000 la gráfica no solo se observaba, sino que se habitaba y generaba de forma autónoma. La representación actual incluye realidad virtual (RV) y realidad aumentada (RA), donde el diseño conduce al usuario a entornos inmersivos. En la etapa más reciente, la representación gráfica arquitectónica se ve profundamente transformada por la incorporación de sistemas basados en IA. A diferencia de herramientas digitales que operaban principalmente como medios de traducción y visualización del proyecto, los modelos computacionales avanzados introducen procesos de generación, evaluación y toma de decisiones automatizadas. Al respecto, Oxman y Oxman (2014), señalan que este cambio implica un desplazamiento desde la representación como mera descripción formal hacia sistemas generativos capaces de producir y optimizar configuraciones espaciales a partir de reglas, datos y algoritmos.

Inteligencia artificial en la representación gráfica actual

Los sistemas digitales han superado su función inicial como simples medios de representación para convertirse en instrumentos activos de diseño. En este nuevo enfoque, las formas ya no se limitan a ser dibujadas, sino que se generan a partir de procesos de cálculo (Fraile, 2018). Entonces, el diseñador puede establecer parámetros, criterios y condiciones específicos como dimensiones y relaciones geométricas para que el sistema digital realice los cálculos y algoritmos y de esta manera producir la geometría, lo que se conoce como diseño paramétrico. Herramientas como *Rhinoceros 3D* con su complemento *Grasshopper* permiten crear formas geométricas complejas y orgánicas utilizando la programación visual fundamentada en la lógica del diseño paramétrico. La plataforma de *Grasshopper* ha permitido el desarrollo de *plugins* relacionados con la IA como *Octopus*, desarrollado en 2012, el cual permite la optimización iterativa de geometrías a partir de objetivos de diseño (Gordillo 2024).

El diseño contemporáneo con el uso de la IA como herramienta, está redefiniendo el flujo de trabajo, permitiendo la automatización de tareas repetitivas que anteriormente se volvían tediosas, y la creación de imágenes complejas mediante análisis de datos masivos. La representación gráfica se convierte en una herramienta cognitiva avanzada (Oxman y Oxman, 2014).

El concepto de IA tiene su origen en la década de 1950 con los trabajos de Alan Turing y John McCarthy y se define como el estudio y desarrollo de sistemas computacionales para realizar funciones que requieren la participación intelectual humana como el aprendizaje, el

razonamiento, la resolución de problemas, la percepción y el uso del lenguaje, centrándose en la imitación del comportamiento humano, en la autonomía y en la adaptabilidad de las máquinas (Gordillo, 2024).

En la arquitectura, la IA se ha introducido en un inicio como una herramienta tecnológica más, con la finalidad de optimizar los procesos repetitivos y gestionar los grandes volúmenes de datos para mejorar la eficiencia productiva. Sin embargo, este enfoque instrumental ha ido evolucionando, para ahora integrarse poco a poco en las fases de creación del diseño a través de múltiples aplicaciones de acceso abierto que permiten la búsqueda más eficiente de información, la generación de ideas, el análisis de datos y principalmente el entrenamiento de IA para el diseño. A partir de lo cual, la IA puede proponer formas, configuraciones espaciales o soluciones a través de imágenes y video en muy poco tiempo, lo que anteriormente era generado paso a paso en diversas plataformas de diseño especializadas que demandaban lapsos prolongados de trabajo, ahora puede orientarse hacia procesos creativos más dinámicos.

La representación gráfica como se ha analizado a lo largo del presente ensayo es la transmisión visual de una idea concebida por el ser humano expresada a través de diferentes medios, entre manuales, digitales e híbridos, tradicionalmente con el control total del ser humano. Con la utilización de la IA generativa, la cual permite crear contenido nuevo basado en el aprendizaje de patrones extraídos de grandes volúmenes de datos, los *prompts* funcionan como una nueva forma de lenguaje de representación, que suple al dibujo y establece parámetros para que el sistema genere múltiples visualizaciones gráficas en diversos formatos. La representación gráfica se convierte en un escenario inmediato en donde el arquitecto no solo plasma lo que imagina, sino que le permite visualizar las posibilidades no previstas, funcionando como un asistente de diseño personal capaz de aprender, generar propuestas y adaptarlas de acuerdo con lo que el diseñador le solicite. La IA puede optimizar procesos al detectar patrones que todavía no han sido identificados. Sin embargo, la planificación de una obra arquitectónica conlleva decisiones que no siempre pueden anticiparse ni resolverse únicamente desde la eficiencia de un *prompt*, pues intervienen factores humanos, económicos, culturales y contextuales (Valencia, 2023).

En este sentido, no toda propuesta generada, aunque sea visualmente impactante o innovadora a simple vista responde a la calidad y a la técnica arquitectónica (Figura 9), este aspecto es uno de los de mayor debate para la profesión en arquitectura, debido principalmente a la utilización desmedida de las representaciones gráficas resultantes de la IA sin

coherencia estructural para su construcción real. Esto apunta a la necesidad de continuar el desarrollo del pensamiento crítico de diseño, de la comprensión profunda de los procesos constructivos y de los principios de representación arquitectónica para así poder valorar y evaluar cada propuesta generada por la IA. Además de sumarse al conjunto de habilidades tradicionales del profesional en arquitectura, la adquisición de la competencia tecnológica en esta herramienta.

Figura 9

Representación gráfica de una propuesta de edificio universitario sustentable e inclusivo a partir de un prompt

Nota. Elaboración propia a partir de *ChatGPT*.

Se piensa que la IA desplazará trabajos y profesiones en poco tiempo, entre ellos al arquitecto, pero solo el arquitecto a través de su formación, experiencia y desarrollo profesional podrá determinar si una propuesta es pertinente, si tiene sentido y coherencia desde criterios estéticos y técnicos. La IA amplia estas capacidades, pero no reemplaza la responsabilidad del arquitecto para comprobar la factibilidad de construcción, el cumplimiento de normativas, las condiciones de seguridad, entre otras actividades que tienen que ver con el criterio técnico y con la responsabilidad ética de un profesional. Como plantea Valencia (2023), las herramientas basadas en la IA no reemplazarán a las personas, pero si empoderarán a quienes sepan utilizarlas. Esto implica una menor demanda de trabajo para ilustradores debido a la creación masiva de imágenes generadas por IA, en contraparte al surgimiento de especializaciones en dar instrucciones (*prompts*) a la IA para la obtención de óptimos resultados en cuanto a representaciones gráficas visuales.

Conclusiones

La evolución de la representación gráfica arquitectónica es un proceso continuo de transformación en el que cada avance tecnológico ha estado íntimamente ligado a cambios culturales, epistemológicos y productivos. Desde la consolidación de los sistemas geométricos y perspectivos en la tradición clásica, pasando por la formalización del dibujo técnico durante la Revolución Industrial y la posterior incorporación de herramientas digitales. La representación gráfica ha funcionado no solo como un medio de comunicación, sino como un instrumento de pensamiento y generación del proyecto arquitectónico.

La geometría descriptiva como lenguaje universal formulado por Monge durante el siglo XVIII, transformó la representación del espacio en un sistema científico basado en proyecciones y relaciones métricas precisas. Ese mismo principio de traducción lógica del mundo tridimensional hoy puede ser el sustento de la inteligencia artificial aplicada a la arquitectura y al diseño, donde algoritmos procesan datos espaciales, generan modelos paramétricos y optimizan soluciones complejas. Históricamente, la IA no rompe con el boceto manual, la maqueta física, el dibujo técnico, o la geometría descriptiva, sino que es una extensión de estos sistemas que se complementan entre sí, convirtiendo el espacio en información estructurada, demostrando que la evolución tecnológica es continuidad metodológica más que sustitución.

De esta forma, la irrupción de la IA no constituye una ruptura aislada, sino la culminación de un largo proceso histórico en el que la representación evoluciona de un sistema descriptivo a uno operativo, generativo, dinámico y flexible. La IA amplía las capacidades de análisis, simulación y producción formal, redefiniendo la relación entre el arquitecto, la técnica y el conocimiento, planteando nuevos retos éticos, culturales y disciplinares que obligan a repensar críticamente el papel de la representación gráfica en la arquitectura contemporánea.

La tecnología digital no sólo ejecuta instrucciones, sino que es partícipe de la exploración, la generación de ideas y alternativas desde las primeras fases del proceso de diseño arquitectónico, entonces el diseño se convierte en un proceso colaborativo entre el diseñador y los sistemas de IA. El diseñador asume un rol de estratega, quien define objetivos, analiza y evalúa resultados para tomar decisiones sobre la propuesta que junto con la IA ha generado. No obstante, la responsabilidad principal sigue siendo humana, el arquitecto establece qué quiere lograr, para quién, en qué entorno y con qué significado social o cul-

tural. Esto a través de la información generada a partir de la IA, por lo que se apunta a una nueva competencia basada en la capacidad de formular instrucciones claras y estructuradas, saber comunicarse con esta nueva herramienta, para transmitir ideas abstractas, criterios espaciales, estéticos y funcionales que pueda procesar para la generación de propuestas coherentes. No se trata solo de producir formas, imágenes y representaciones gráficas al mayoreo, sino de gestionar y evaluar su aplicabilidad.

La IA actualmente se está convirtiendo en una herramienta importante para la gráfica arquitectónica en donde el desarrollo del pensamiento crítico juega un papel crucial. Se busca equilibrar las distintas herramientas de representación gráfica y la adquisición de competencias en cada una, para lograr un proyecto consistente tanto en diseño como en el aspecto técnico. El dominio del dibujo técnico, la geometría, el modelado digital y las herramientas basadas en IA son recursos complementarios donde cada uno aporta características específicas al proceso de diseño y construcción.

El enfoque de la representación gráfica en tiempos de la IA debe prevalecer en su función para explorar, comprender y comunicar el proyecto arquitectónico y no solo reducirse a la generación de imágenes sin criterio. Por tal razón, la IA depende del juicio crítico del arquitecto para validar su resultado. Esta herramienta no sustituye al profesionista, se convierte en un recurso valioso que apoya su quehacer, pero carece de empatía y de comprensión humana. En este sentido, el verdadero creador que abre paso al futuro de las ciudades es el ser humano, donde sus necesidades van evolucionando, pero las básicas siempre prevalecen. Diseñar, construir y habitar los espacios en tiempos de la IA es emocionante y a la vez inquietante, esta herramienta emerge como un sistema que puede vincular pensamiento complejo de diseño, tecnología y cultura bajo una estructura de responsabilidad y ética humana.

Referencias

Carpo, M. (2001). *Architecture in the age of printing: Orality, writing, typography, and printed images in the history of architectural theory*. MIT Press.

Cutieru, A. (2021, 12 de abril). *Pioneros: 6 prácticas que llevan la Inteligencia Artificial a la arquitectura* [Pioneers: 6 Practices Bringing AI into Architecture]. ArchDaily México. (Trad. Baraya, Santiago) https://www.archdaily.mx/mx/937256/pioneros-6-practicas-que-llevan-la-inteligencia-artificial-a-la-arquitectura

Gelb, I. J. (1987). *Historia de la escritura*. Alianza Editorial.

Fraile, M. A. (2018). *Disrupción digital: el comienzo de una nueva revolución*. TRP21.

Gombrich, E. H. (1986). *Arte e ilusão: Um estudo da psicologia da representação pictórica*. Martins Fontes.

Gombrich, E. H. (1995). *Historia del arte*. Alianza.

Gordillo Díaz, A. (2024). El papel de la inteligencia artificial en la producción del diseño: Colaboración o competencia. *Bitácora Arquitectura*, *1*(54), 70–81. https://doi.org/10.22201/fa.14058901p.2024.54.89983

Hauser, A. (1972). *História social da literatura e da arte* (Cap. 1). Mestre Jou.

Hirsch, R. (2017). *Seizing the light: A history of photography* (3rd ed.). Routledge.

Honnecourt, V. (ca. 1230). Folio del Carnet: Fachada gótica [Ilustración]. *Wikimedia Commons*. https://commons.wikimedia.org/

Kolarevic, B. (2003). *Architecture in the digital age: Design and manufacturing*. Spon Press.

Martínez, C. (2025). La arquitectura virtual con inteligencia artificial. *Asinea* (60), 84–89. https://asinea.mx/revista-asinea

Mendoza, R., Campos, E., & Villarreal, L. (2025). Germán Benítez:diseño, ambiente y cultura de la vivienda de ciudad. En Y. Díaz, V. Martínez y M. Tostado et al. (Coord). *Ciencias de la Arquitectura y Desarrollo Territorial* (pp. 49-74). Editorial Racionalidades.

Monge, G. (1799). *Géométrie descriptive*. Baudouin.

Negroponte, N. (1970). *The architecture machine*. MIT Press.

Oxman, R., & Oxman, R. (2014). *Theories of the digital in architecture*. Routledge.

Pagnotta, B. (2015, 25 de marzo). *Clásicos de Arquitectura: Museo Guggenheim Bilbao / Frank Gehry"* [AD Classics: The Guggenheim Museum Bilbao / Gehry Partners]. ArchDaily México. (Trad. Yunis, Natalia). https://www.archdaily.mx/mx/764294/clasicos-de-arquitectura-museo-guggenheim-bilbao-frank-gehry

Palladio, A. (1570). Interior of San Paolo fuori le Mura, Rome [Dibujo]. *Wikimedia Commons*. https://www.1st-art-gallery.com/media/catalog/product/cache/a67918ceb14cd34227a9a3ea0f98d29a/o/r/original_13038.webp

Panofsky, E. (1991). *Significado en las artes visuales*. Perspectiva.

Rameessos. (2015, enero 6). *Cave painting in the Altamira Cave*. World History Encyclopedia. https://www.worldhistory.org/image/3537/cave-painting-in-the-altamira-cave/

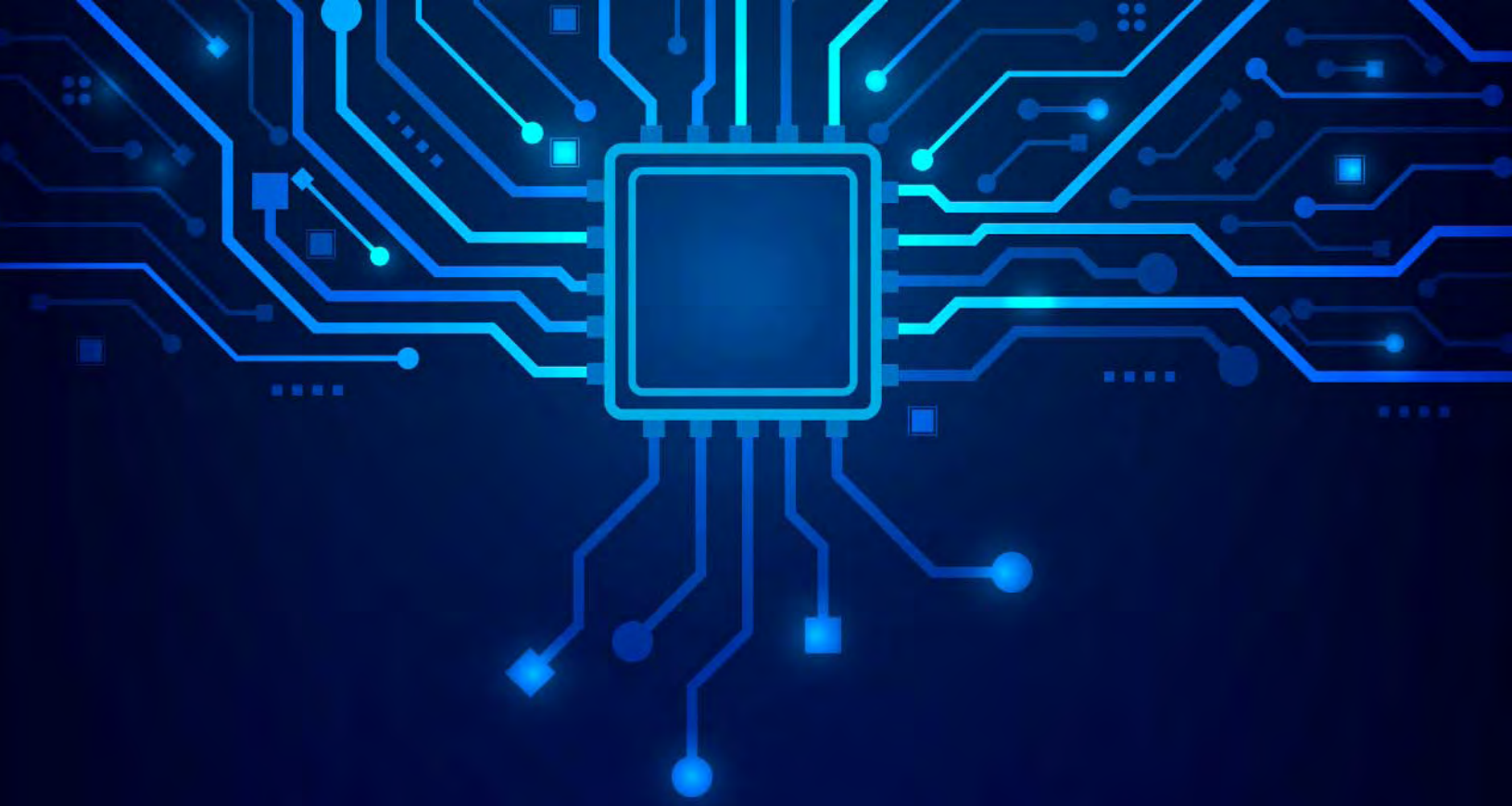
Schwartz, G. M. (2021). Tell Umm el-Marra (Syria) and Early Alphabetic in the Third Millennium: Four Inscribed Clay Cylinders as a Potential Game Changer. *Pasiphäe*, *XIII*, 255-266

Sketchup. (2026, 03 de febrero). https://sketchup.mx/

Valencia, N. (2023, 06 de diciembre). El impacto de la inteligencia artificial en la arquitectura en 2024 (y más allá) [The Impact of AI Tools on Architecture in 2024 (and Beyond)]. *ArchDaily México*. https://www.archdaily.mx/mx/1010103/el-impacto-de-la-inteligencia-artificial-en-la-arquitectura-en-2024-y-mas-alla?ad_source=search&ad_medium=projects_tab&ad_source=search&ad_medium=search_result_all

Villanueva Bartrina, L. (2005). *Historia de la representación arquitectónica y del diseño* (Vol. 1). Universitat Politécnica de Catalunya.

Ware, M. (1999). *Cyanotype: The history, science and art of photographic printing in Prussian blue*. Science Museum



18.

Conocimiento o apariencia: el estatuto epistémico de la inteligencia artificial generativa

Tomás Peralta Palazón
Universidad de Barcelona, España
https://orcid.org/0009-0006-7924-325X

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.18

18.

Conocimiento o apariencia: el estatuto epistémico de la inteligencia artificial generativa

Introducción

La Inteligencia Artificial Generativa (IAG) ha intensificado su presencia pública en los últimos años y se ha convertido, posiblemente, en un fenómeno mediático. Su expansión no se explica solo por el avance técnico, sino también por la lógica algorítmica de la viralización digital. Las múltiples aplicaciones que ofrece, especialmente en la generación de imágenes, vídeos y textos, han favorecido una circulación masiva de contenidos que se reproducen, se transforman y se amplifican con gran rapidez y que han ido integrándose de manera casi natural en la vida cotidiana.

En este sentido, esta mayor visibilidad ha ido acompañada de la proliferación de contenidos generados artificialmente en los que, en muchos casos, resulta cada vez más difícil distinguir lo real de lo irreal. Imágenes hiperrealistas o vídeos sintéticos que no proceden de una grabación directa de la realidad se consumen a menudo sin que se cuestione su origen ni su estatuto de verdad. La creciente dificultad para distinguir lo real y lo irreal no remite únicamente a un problema de circulación de contenidos, sino que podría estar señalando una transformación más profunda en los criterios de validación y de autoridad del conocimiento (Floridi, 2014). Por ello, conviene detenerse en las implicaciones que esta integración tecnológica introduce en nuestras formas de comprender y legitimar el conocimiento.

Llegados a este punto, cabe proponer una reflexión en torno a la siguiente cuestión: ¿estamos ante una burbuja tecnológica, una revolución técnica o ante un posible cambio de paradigma? Bajo esta primera coyuntura, pensar la IAG como una burbuja supondría atender a su rápida expansión, al peso de la expectación y a la lógica de la aceleración, dejando abierta la posibilidad de que su centralidad actual no sea necesariamente estable ni duradera. Por su parte, interpretarla como revolución técnica implicaría situarla en la historia del desarrollo tecnológico como una innovación capaz de transformar procesos,

reorganizar prácticas y redefinir competencias en distintos ámbitos de la sociedad, sin que ello suponga necesariamente una alteración de los marcos desde los que se comprende y legitima el conocimiento.

Sin embargo, una interpretación de mayor alcance sería entenderla como un posible cambio de paradigma. Esto supondría comprender el fenómeno como el inicio de una transformación más profunda en la que se verían afectados los marcos desde los cuales se organiza, produce y legitima el conocimiento. Bajo esta perspectiva, un cambio de paradigma implicaría una transformación de los supuestos, los modelos y los métodos que orientan la comprensión del mundo y que sirven de base para la resolución de problemas y el avance del conocimiento (Kuhn, 1971).

Sobre esta cuestión pueden distinguirse distintas posiciones. Entre ellas, cabría sostener que las tres posibilidades podrían coexistir sin excluirse mutuamente. Esto no implicaría una contradicción entre ellas, sino que permitiría entenderlas como distintos niveles de interpretación de un mismo fenómeno: la expansión propia de una burbuja tecnológica, la transformación característica de una revolución técnica y los indicios de una reconfiguración más profunda de los marcos de producción y legitimación del conocimiento. Desde esta mirada, dicha coexistencia permitiría comprender la complejidad del momento sin reducirlo a una única categoría explicativa. No obstante, será, en última instancia, el tiempo quien permita valorar el alcance de esta transformación.

Sin embargo, esta discusión tipológica admite más matices que enriquecen la comprensión del fenómeno. Más allá de su dimensión histórica, social o tecnológica, la IAG plantea también interrogantes filosóficos y epistemológicos, que constituyen el eje del presente capítulo. En este contexto, la apariencia de realidad que acompañan a muchos contenidos generados por sistemas de inteligencia artificial empieza a ganar terreno en la forma en que esos contenidos se interpretan y aceptan, desplazando la atención desde su estatuto de verdad hacia su plausibilidad formal. Lo que parece real tiende a aceptarse como tal, incluso cuando las condiciones de su producción quedan en un segundo plano. En la práctica educativa, esto se traduce en textos formalmente impecables que, al contrastarse, revelan lagunas conceptuales, citas inverificables o inferencias no justificadas. En este escenario, la cuestión remite directamente a la filosofía: qué implica para nuestra comprensión del conocimiento que la coherencia, la corrección formal y la verosimilitud puedan generarse sin mediación de comprensión humana.

Desde una perspectiva filosófica, la capacidad de la IAG para producir contenidos formalmente convincentes introduce una tensión entre conocimiento y apariencia. Aquello que tradicionalmente se asociaba al saber, la coherencia argumentativa, la estructura racional o la adecuación estilística, puede hoy reproducirse sin que medie necesariamente justificación ni responsabilidad epistémica por parte de un sujeto cognoscente. Esta situación favorece la confusión entre conocer y aparentar conocer y obliga a revisar los criterios mediante los cuales se atribuye valor epistémico a la información que circula.

Este desplazamiento abre interrogantes centrales: ¿qué significa conocer en un contexto en el que determinadas operaciones cognitivas pueden externalizarse, en un sentido cercano a lo planteado por Clark y Chalmers (1998)? ¿En qué se distingue la comprensión de la mera elaboración formal de un contenido correcto cuando ambos ofrecen resultados aparentemente equivalentes? *¿Qué lugar ocupan el juicio humano y la agencia humana cuando la generación automática de textos, imágenes o argumentos se integra sin fricción en nuestras prácticas cotidianas?* La dificultad para responder a estas cuestiones pondría de manifiesto que el problema no sería meramente técnico ni comunicativo, sino propiamente filosófico y epistemológico.

Conocer y comprender: fundamentos epistemológicos

La distinción entre el conocimiento y la mera opinión, así como el interrogante sobre la naturaleza misma del acto de conocer, constituyen uno de los ejes de la tradición epistemológica que trasciende de la simple acumulación de información. En el diálogo *Teeteto*, Platón sostiene que el conocimiento consiste en que una creencia verdadera debe ser sustentada por una justificación racional. Con ello sugiere que no basta con que algo sea verdadero, sino que debe fundarse en razones que permitan reconocerlo como tal. De este modo, se configura la conocida relación entre verdad, creencia y justificación (Platón, 2021). Esta caracterización no se describe únicamente un estado psicológico, sino que incorpora una dimensión normativa, en la medida en que exige fundamentación racional para aquello que se presenta como verdadero.

Este planteamiento también se proyecta en la alegoría de la caverna (Platón, 2020), donde la toma de las sombras por la realidad no responde a un error voluntario sino a una limitación estructural en el acceso a lo real. Bajo esta premisa, el tránsito de la apariencia a la verdad y de la opinión al conocimiento fundado en principios inteligibles exige someter la experiencia al examen racional para superar la inmediatez de lo sensible.

En el contexto actual, la IAG reabre, bajo condiciones técnicas inéditas, la cuestión de la relación entre apariencia y verdad. Los textos, imágenes o argumentos que produce pueden presentarse con un grado de coherencia y verosimilitud que favorece su aceptación sin un contraste suficiente. Como en la caverna, lo que se ofrece ante nosotros posee forma, estructura y consistencia, hasta el punto de resultar difícilmente cuestionable. Sin embargo, ello no garantiza que constituya conocimiento en sentido estricto. En este escenario, estos sistemas tensionan la distinción entre conocimiento y mera apariencia de conocimiento, introduciendo un desafío que no es solo estético, en el sentido de la forma persuasiva y convincente de los contenidos, sino también ético y académico.

Si aquello que tradicionalmente identificamos como signos de conocimiento puede generarse sin comprensión ni responsabilidad intelectual, resulta necesario revisar los criterios que regulan su uso en el ámbito universitario. El empleo de estas herramientas, como *ChatGPT*, *GitHub, Copilot* o *Gemini*, plantea retos en el ámbito de la ética tecnoeducativa, tales como el posible incremento del plagio, la reproducción de sesgos, la dependencia tecnológica o la erosión del pensamiento crítico (Peralta Palazón et al., 2025).

Estos retos pueden entenderse como nuevas *sombras*, y la cuestión no radica tanto en su existencia como en la forma en que se afrontan. En el ámbito académico, el tránsito hacia el conocimiento no pasa por rechazar la herramienta, sino por integrarla desde criterios de responsabilidad e integridad académica. Ello exige someter sus resultados a contraste, justificar aquello que se adopta y asumir la autoría intelectual de lo que se afirma. La integridad académica no aparece así como un requisito formal, sino como una condición que orienta la búsqueda del conocimiento más allá de la mera apariencia. Esta exigencia remite, en último término, a la pregunta por qué entendemos por conocimiento y bajo qué condiciones puede considerarse legítimamente fundado.

Con todo, la definición clásica del conocimiento como creencia verdadera justificada fue objeto de discusión en la filosofía contemporánea. Gettier (1963) mostró que pueden darse situaciones en las que una creencia es verdadera y está justificada y, sin embargo, no constituir conocimiento en sentido estricto. Estos casos dan lugar a un replanteamiento de las condiciones bajo las cuales una afirmación puede considerarse epistémicamente válida. En consecuencia, la atención deja entonces de centrarse exclusivamente en la estructura proposicional de la creencia para incorporar también el análisis de la fiabilidad de los procesos que la generan y de la responsabilidad que implica su formulación.

En este contexto, y más allá de la mera corrección proposicional, la noción de comprensión adquiere una relevancia particular. Comprender ya no equivaldría simplemente a poseer enunciados correctos, sino que también implicaría captar las relaciones internas que articulan un contenido dentro de un marco explicativo más amplio. Como ha señalado Kvanvig (2003), la comprensión supone una integración estructural que va más allá de la acumulación de proposiciones verdaderas. Mientras el conocimiento puede referirse a afirmaciones concretas, la comprensión remite a la capacidad de situarlas en un entramado coherente de razones y conexiones.

En este marco, la distinción entre conocimiento y mera corrección adquiere especial relevancia cuando se introduce la dimensión de la responsabilidad epistémica. Desde la epistemología de las virtudes, se ha defendido que conocer no consiste únicamente en acertar, sino en hacerlo por las razones adecuadas y mediante disposiciones intelectuales apropiadas, como la honestidad, la atención o el rigor (Sosa, 2007; Zagzebski, 1996). El valor epistémico no se reduce al resultado correcto, sino que depende del modo en que este se obtiene. La justificación no es, por tanto, un añadido externo, sino parte constitutiva del acto mismo de conocer.

A la luz de este planteamiento puede situarse con mayor precisión el problema que plantea la IAG. Si determinados sistemas son capaces de producir contenidos formalmente coherentes, plausibles e incluso argumentativamente sofisticados sin mediación de comprensión humana, la frontera entre conocer y aparentar conocer se vuelve menos nítida. La cuestión ya no reside solo en la veracidad puntual de un enunciado, sino en los criterios mediante los cuales atribuimos autoridad, fiabilidad y legitimidad epistémica en un entorno informacional transformado (Floridi, 2011, 2014). En una realidad progresivamente estructurada por la infosfera, entendida como el entorno global de información en el que interactúan agentes humanos y artificiales, no cambian únicamente las condiciones de circulación del conocimiento, sino también los marcos desde los que evaluamos su origen, su justificación y su valor. Esto obliga a reconsiderar el lugar que ocupa el sujeto en estos procesos.

La mediación técnica se convierte así en una cuestión central. Tradicionalmente, la epistemología ha asumido que el sujeto cognoscente constituye el centro desde el cual se articulan las operaciones cognitivas. La tesis de la mente extendida sostiene, sin embargo, que ciertos procesos pueden configurarse en interacción con artefactos externos, formando sistemas acoplados que participan funcionalmente en la actividad cognitiva (Clark y Chal-

mers, 1998). Esta hipótesis no elimina la agencia humana, pero sí problematiza los límites entre interioridad y soporte técnico.

La mediación tecnológica no puede entenderse únicamente como instrumento neutro, sino como un elemento que influye en la organización y presentación del conocimiento. Incluso sin adoptar posiciones radicales, es preciso reconocer que los dispositivos técnicos configuran el acceso a la información y condicionan los criterios de evaluación. En el caso de la IAG, esta mediación adquiere un carácter cualitativamente distinto, puesto que el sistema no solo almacena o transmite información, sino que produce contenidos con apariencia de autonomía cognitiva.

La cuestión decisiva no es, por tanto, si la producción técnica resulta eficiente o útil, sino si cumple las condiciones normativas asociadas al saber y al conocimiento. Aunque determinadas operaciones puedan externalizarse, la exigencia de justificación y responsabilidad no desaparece, ya que toda atribución de conocimiento implica un marco de imputación ética. El problema no es meramente funcional, sino estructural, en la medida en que afecta a la distinción entre información, comprensión y conocimiento. El interrogante se vuelve más agudo cuando los rasgos formales del conocimiento pueden reproducirse sin que exista un sujeto que asuma la responsabilidad de lo que se afirma.

Inteligencia artificial generativa, apariencia epistémica y autonomía cognitiva

En el marco de la IAG, la relación entre producción discursiva, apariencia epistémica y autonomía cognitiva exige precisar conceptualmente estos términos, pues la producción de un discurso coherente no equivale a conocimiento. Los modelos actuales de lenguaje se fundamentan en la arquitectura *Transformer*, introducida por Vaswani et al. (2017), que sustituye las estructuras recurrentes por mecanismos de atención capaces de modelar dependencias contextuales a lo largo de secuencias completas. Sobre esta base, el escalado masivo de parámetros y datos de entrenamiento ha permitido desarrollar modelos de gran tamaño con elevada capacidad de generalización lingüística, como muestran Brown et al. (2020) en el caso de *GPT-3*. No obstante, esta sofisticación técnica se inscribe en un marco estrictamente funcional: se trata de optimizar la predicción probabilística de secuencias lingüísticas a partir de patrones estadísticos aprendidos, no de fundamentar la existencia de comprensión semántica o intencionalidad. La reproducción de rasgos formales asociados a la comprensión no implica, por sí misma, la presencia de significado entendido.

Esta distinción adquiere relevancia filosófica si se atiende al argumento desarrollado por John Searle (1980) en *Minds, Brains, and Programs*. La manipulación sintáctica de símbolos conforme a reglas formales no equivale a comprensión. El experimento mental de la habitación china pretende mostrar que un sistema puede generar respuestas correctas en virtud de reglas operativas sin entender el significado de los símbolos que procesa. Searle no niega la eficacia causal de los sistemas computacionales, sino la identificación entre ejecución algorítmica y estados mentales intencionales. Aplicado al ámbito de la IAG, esto implica que la coherencia textual y la adecuación contextual no constituyen, por sí mismas, actos de conocimiento. La forma del razonamiento puede estar presente sin que exista un sujeto que comprenda aquello que articula.

Desde la perspectiva del usuario, la respuesta generada por estos modelos puede presentarse con apariencia de deliberación racional, como si fuera fruto de un proceso consciente de reflexión. El discurso organiza premisas, desarrolla inferencias y concluye con afirmaciones estructuradas. Aquí emerge una tensión clásica de la epistemología. El análisis del conocimiento como creencia verdadera justificada fue problematizado por Edmund Gettier (1963), quien mostró que una creencia puede ser verdadera y estar justificada sin constituir conocimiento. El debate posterior evidenció que la mera adecuación formal o la presencia de razones aparentes no garantizan el estatuto epistémico pleno. En el contexto de la IAG, la corrección gramatical, la coherencia lógica y la plausibilidad argumentativa pueden inducir a aceptar un contenido como conocimiento cuando carece de verificación rigurosa o contiene inexactitudes factuales. La apariencia epistémica se refuerza precisamente porque el discurso adopta la forma externa de la argumentación racional.

En el marco de la infosfera, este fenómeno adquiere una dimensión más amplia. Como sostiene Luciano Floridi (2011), la vida contemporánea se desarrolla en un entorno informacional mediado por infraestructuras tecnológicas complejas. La autoridad epistémica, es decir, los criterios mediante los cuales atribuimos validez al conocimiento, se reconfigura cuando el acceso a este queda mediado por sistemas algorítmicos que no solo seleccionan información, sino que contribuyen a estructurarla. Entre estos sistemas, la IAG no se limita a distribuir contenidos, sino que interviene en su configuración formal. Esta mediación puede generar una impresión de competencia y exhaustividad que refuerza la confianza del usuario en resultados cuyo proceso de elaboración no es transparente y que pueden incorporar sesgos difíciles de detectar, especialmente cuando parte de la organización del discurso ha sido delegada al sistema.

La externalización cognitiva no constituye una novedad histórica. Bernard Stiegler (1994) describió la técnica como un proceso de exteriorización de la memoria y de las capacidades humanas. La escritura y otros dispositivos técnicos han funcionado como soportes en los que la experiencia y el saber se sedimentan más allá del sujeto. Sin embargo, la IAG introduce un desplazamiento cualitativo al intervenir en la estructuración misma del discurso. Cuando el sujeto delega en el sistema la organización de un razonamiento, la formulación de hipótesis o la redacción de un análisis, parte del proceso de articulación intelectual se desplaza hacia el dispositivo técnico. Esta delegación puede resultar útil, pero no solo modifica la relación del sujeto con su ejercicio reflexivo, sino que incide también en la configuración misma de la comprensión.

El estatuto del sujeto cognoscente en entornos algorítmicos queda así tensionado. Hannah Arendt (1971) advirtió que no cabe confundir el funcionamiento técnico con la actividad de pensar, que implica examen crítico y asunción de responsabilidad. Pensar supone sostener un diálogo interior que somete a prueba las propias afirmaciones. Si la producción automática sustituye el proceso de elaboración personal sin mediación crítica, el riesgo no reside solo en el error puntual, sino en el debilitamiento progresivo del juicio. La eficiencia técnica puede ocultar una heteronomía epistémica creciente, es decir, una dependencia del juicio respecto del sistema cuando la respuesta se adopta sin contraste ni apropiación comprensiva.

Desde una perspectiva normativa, esta cuestión se vincula con la autonomía cognitiva. La autonomía implica asumir como propias las razones que fundamentan una afirmación tras un proceso de examen. Cuando la autoridad se desplaza hacia la respuesta generada por un sistema, la mediación tecnológica puede convertirse en instancia decisiva. En este sentido, la UNESCO (2022), en su Recomendación sobre la ética de la inteligencia artificial, adopta un enfoque humanocéntrico que sitúa la dignidad humana como criterio rector y subraya la necesidad de preservar una supervisión humana significativa. La advertencia es clara: la responsabilidad no puede diluirse en el funcionamiento automatizado ni la agencia humana quedar sustituida por decisiones algorítmicas. En el ámbito académico, esta exigencia resulta especialmente relevante, ya que el conocimiento no se reduce a la producción de resultados correctos, sino que implica formación del juicio y responsabilidad intelectual.

La naturalización de la IAG entraña riesgos epistemológicos cuando la apariencia de coherencia sustituye el contraste crítico. La corrección formal y la plausibilidad discursiva pueden producir apariencia de conocimiento sin que medie comprensión suficiente. La ex-

ternalización puede convertirse en delegación del juicio si no se acompaña de un ejercicio consciente de apropiación. La autonomía se compromete cuando la mediación tecnológica se asume como neutral y autosuficiente. Mantener la primacía del examen racional y de la responsabilidad intelectual constituye, en este escenario, una exigencia formativa indispensable para preservar la distinción entre apariencia y saber fundado.

Un ejemplo de ello puede encontrarse en el ámbito universitario, cuando un estudiante presenta un trabajo elaborado con apoyo de un sistema de IAG que muestra coherencia argumentativa y corrección formal. El texto puede estar bien estructurado y ofrecer definiciones y argumentos plausibles. Sin embargo, si el trabajo tuviera que ser defendido o explicado, podrían evidenciarse dificultades para justificar las afirmaciones formuladas o aclarar las conexiones conceptuales que el propio texto desarrolla, lo que pondría de manifiesto una cierta desvinculación intelectual y omisión reflexiva respecto del proceso mediante el cual fue elaborado.

Por tanto, aunque el trabajo cumple aparentemente con los criterios formales de un ejercicio académico, esto no supone necesariamente comprensión efectiva ni apropiación intelectual. Esta situación, podría comprometer los propios procesos de enseñanza y aprendizaje, en la medida en la que formación del juicio y al desarrollo intelectual del estudiante se ve afectada.

Discusión y conclusiones

La IAG no puede reducirse a una innovación técnica susceptible de describirse únicamente en términos de eficiencia predictiva o modelado estadístico del lenguaje. El núcleo del problema no reside en su arquitectura computacional, sino en el estatuto epistémico que adquieren sus productos cuando se integran en prácticas académicas y sociales. Lo decisivo no es solo qué generan estos sistemas, sino bajo qué condiciones aquello que generan es interpretado como conocimiento legítimo. Desde una perspectiva histórica, la aparición de nuevas tecnologías y métodos ha influido en distintas épocas en las condiciones de producción y transmisión del saber, como ocurrió, entre otras, con la imprenta, la consolidación del método científico moderno o la digitalización de la información.

Desde la tradición epistemológica, conocer ha implicado algo más que formular enunciados correctos. La justificación, la integración comprensiva de razones y la asunción de responsabilidad intelectual constituyen dimensiones estructurales del saber. En este con-

texto, la capacidad de la IAG para reproducir la forma externa del razonamiento reabre, bajo condiciones técnicas inéditas, la tensión entre apariencia y conocimiento. La coherencia argumentativa y la adecuación estilística pueden simular los rasgos del saber sin que medie comprensión ni imputación de razones. La corrección formal no basta para hablar de fundamentación. La validez epistémica exige consistencia lógica y justificación adecuada.

Esta constatación nos exige situar el análisis en un plano normativo. Desde una perspectiva de ética tecnoeducativa, la cuestión no consiste en aceptar o rechazar la herramienta, sino en determinar los criterios que deben orientar su integración responsable en contextos formativos y sociales. La tecnología no es axiológicamente neutra cuando interviene en la producción y organización del discurso. Su incorporación transforma prácticas, hábitos cognitivos y modos de validación. Por ello, la discusión no puede limitarse a la dimensión funcional, sino que debe atender a las condiciones bajo las cuales se preserva la autonomía cognitiva y la responsabilidad epistémica del sujeto.

En el ámbito académico, esta exigencia se articula directamente con el principio de integridad académica. La integridad no se limita a evitar el plagio o la falsificación de fuentes, sino que se fundamenta en principios como la honestidad, la confianza, el respeto, la justicia y la responsabilidad, que orientan tanto el comportamiento del estudiante como su compromiso con el conocimiento (Vital Carrillo, 2023). Promover una cultura de integridad académica implica, por tanto, no solo establecer normas, sino educar en valores que favorezcan una ciudadanía crítica y responsable. Cuando la estructuración del razonamiento o la redacción argumentativa se delegan en sistemas algorítmicos sin apropiación comprensiva, el riesgo no es únicamente normativo, sino también formativo, pues afecta a la configuración misma del juicio.

Los marcos internacionales recientes subrayan asimismo la necesidad de mantener una supervisión humana significativa en el uso de sistemas de inteligencia artificial. La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura ([UNESCO], 2021) sitúa la dignidad humana, la agencia y la responsabilidad en el centro de cualquier implementación tecnológica. En la misma línea, la European Commission (2021), en su propuesta de regulación sobre inteligencia artificial, insiste en la importancia de la transparencia y la rendición de cuentas. Estos documentos no resuelven el problema epistemológico, pero evidencian que la imputación de responsabilidades no puede diluirse en la eficiencia algorítmica.

El análisis desarrollado a lo largo de este capítulo ha sido deliberadamente conceptual y no pretende clausurar el debate. La evolución acelerada de estos sistemas aconseja prudencia en la formulación de juicios definitivos y pone de relieve la necesidad de profundizar en el diálogo entre análisis filosófico y evaluación empírica. Se abren así interrogantes que requieren ulterior desarrollo. ¿Cómo se reconfigura la autoridad epistémica cuando parte de la producción discursiva es generada por sistemas no humanos? ¿En qué condiciones puede la externalización cognitiva resultar compatible con la autonomía intelectual y la formación del juicio crítico?

La IAG no elimina la exigencia de justificación ni la responsabilidad intelectual. Al contrario, las vuelve más patentes. En un contexto en el que la forma externa del saber puede reproducirse con notable fidelidad técnica, la distinción entre apariencia y conocimiento adquiere una nueva dimensión. Por ello, la cuestión decisiva no radica en la capacidad de estos sistemas para producir discursos plausibles o formalmente coherentes, sino en si se preservan las condiciones normativas que hacen posible el conocimiento mismo. En última instancia, lo que está en juego es la preservación de las condiciones que hacen posible el conocimiento.

Referencias

Arendt, H. (1971). *The life of the mind*. Harcourt Brace Jovanovich.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language models are few-shot learners. Advances in Neural Information Processing Systems*, 33, 1877–1901.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

European Commission. (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence* (*Artificial Intelligence Act*) *and amending certain Union legislative acts* (*COM*(*2021*) *206 final*). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Floridi, L. (2011). *The philosophy of information.* Oxford University Press.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123. https://doi.org/10.1093/analys/23.6.121

Kuhn, T. S. (1971). *La estructura de las revoluciones científicas* (A. Contín, Trad.). Fondo de Cultura Económica. (Trabajo original publicado en 1962)

Kvanvig, J. L. (2003). The value of knowledge and the pursuit of understanding. Cambridge University Press. https://doi.org/10.1017/CBO9780511498909

Peralta Palazón, T., Galván Fernández, C., & Buxarrais, M. R. (2025). Inteligencia artificial generativa, ética tecnoeducativa e integridad académica en la educación superior. En *Proceedings of the International Conference on Educational Technology and Computers* (*ICETC 2025*) (pp. 564–569). IEEE. https://doi.org/10.1109/ICETC66579.2025.11387569

Platón. (2020). *República* (C. Eggers Lan, Trad.). Gredos. (Obra original del siglo IV a. C.)

Platón. (2021). Teeteto (Á. Vallejo Campos, Trad.). En C. García Gual (Ed.), *Diálogos V: Parménides, Teeteto, Sofista, Político* (Biblioteca Clásica Gredos). Gredos. (Obra original del siglo IV a. C.)

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1). Oxford University Press.

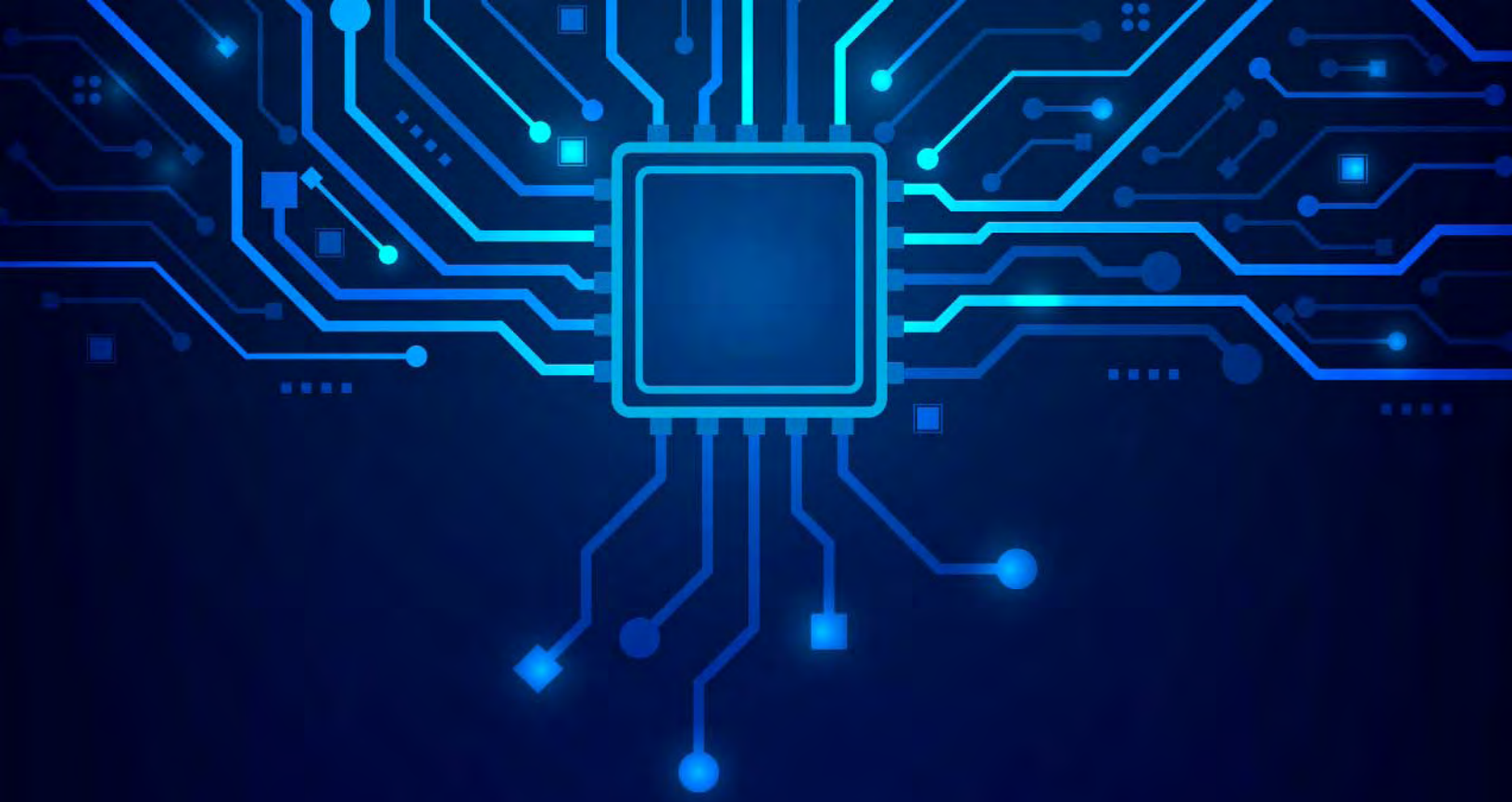
Stiegler, B. (1994). *La technique et le temps 1: La faute d'Épiméthée*. Galilée.

UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762

Vital Carrillo, M. (2023). Integridad académica. *Vida Científica: Boletín Científico de la Escuela Preparatoria No. 4*, *11*(22), 42–44. https://doi.org/10.29057/prepa4.v11i22.11000

Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.



19.

Entre el prompt y el diseño: experiencias de co-creación crítica humano-inteligencia artificial en educación superior

Juan Carlos Lobato-Valdespino
Universidad Michoacana de San Nicolás de Hidalgo, México
ORCID: 0000-0001-6173-3847

Claudia Margarita García Paulín
Universidad Michoacana de San Nicolás de Hidalgo, México
ORCID: 0000-0002-0641-8008

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.19

19.

Entre el prompt y el diseño: experiencias de co-creación crítica humano-inteligencia artificial en educación superior

Introducción

«La inteligencia artificial amplifica mis sesgos tanto como mis posibilidades; trabajar con ella es dialogar con mi propia sombra.» Esa frase apareció en la bitácora de una estudiante de arquitectura a mitad de un taller intensivo. Nadie le pidió que reflexionara sobre sesgos algorítmicos. Después de sesenta iteraciones con un generador de imágenes, algo cambió en su forma de mirar lo que la máquina le devolvía. Este capítulo busca documentar y comprender ese giro.

La educación superior aún no sabe bien qué hacer con la inteligencia artificial generativa (IAG). Las instituciones oscilan entre la prohibición y la adopción entusiasta, y mientras tanto los estudiantes ya están usando *Midjourney*, *DALL-E* o *Stable Diffusion* para resolver en segundos tareas que antes tomaban semanas de bocetaje manual. En disciplinas proyectuales —arquitectura, diseño industrial, artes visuales— el impacto es especialmente agudo porque la generación de imágenes toca el núcleo mismo del proceso creativo (Leach, 2022). La velocidad seduce. Pero también oculta un problema de fondo.

Ese problema se manifiesta como lo que esta investigación denomina síndrome de la instantaneidad: la prioridad del resultado visual inmediato sobre la reflexión conceptual, la viabilidad constructiva y la responsabilidad ética. No es una metáfora. En los talleres que aquí se describen, se observó un patrón recurrente: ante la posibilidad de producir decenas de imágenes por hora, los estudiantes tendían a elegir las más vistosas sin preguntarse qué significaban, a quién representaban o qué supuestos culturales arrastraban desde los datos de entrenamiento. Chen et al. (2024) y Akolekar et al. (2025) han documentado dinámicas similares en otros contextos formativos.

El riesgo va más allá de lo estético. Los modelos de generación visual tienden a reproducir de manera desproporcionada estéticas occidentales, tipologías habitacionales convencionales y representaciones demográficamente homogéneas. Crawford (2021) ha mapeado con precisión cómo las arquitecturas de datos cristalizan desigualdades preexistentes. Mohamed et al. (2020) formularon el problema en términos de colonialidad incrustada en los sistemas de IA. Cuando se trata de formar profesionales que diseñarán los espacios donde otros habitarán, aceptar esas sugerencias sin filtro crítico equivale a delegar decisiones de diseño en sistemas que no rinden cuentas.

Frente a este panorama, la investigación que aquí se presenta propone una respuesta pedagógica específica: utilizar el *storytelling* —la construcción de narrativas especulativas— como andamiaje para la interacción crítica con la IA. La narrativa no es aquí un ornamento. Bruner (2003) argumentó que constituye un modo de pensamiento tan legítimo como el lógico-formal, una forma primaria de organizar la experiencia y de otorgar sentido a la acción. En el contexto del diseño, narrar antes de generar imágenes significa establecer intenciones, inventar personajes que habitarán los espacios, imaginar conflictos que la arquitectura deberá negociar. Ese acto de intencionalidad reemplaza el ensayo-error aleatorio por una búsqueda con dirección.

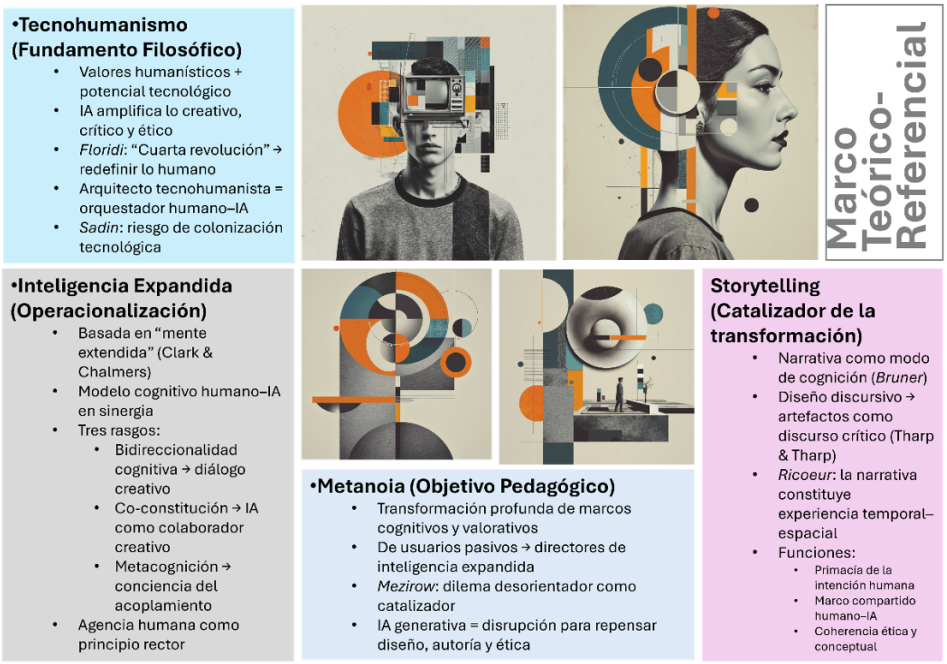
El marco teórico se articula en torno a la noción de que la cognición no ocurre exclusivamente dentro del cerebro, sino que se distribuye entre cuerpo, herramientas y entorno. Clark y Chalmers (1998) formularon esta tesis en su influyente ensayo sobre la mente extendida; Paul (2021) la desarrolló para contextos de aprendizaje cotidiano. La implicación pedagógica es directa: un sistema humano-IA puede funcionar como aparato cognitivo integrado, pero solo si la interacción está mediada por un propósito crítico. La post-fenomenología añade un matiz indispensable: las tecnologías no son instrumentos neutros que el sujeto simplemente usa, sino mediadores que configuran activamente lo que percibimos y lo que podemos imaginar (Verbeek, 2015). Si la IA configura la imaginación del estudiante, la pregunta pedagógica pertinente no es cómo usar la herramienta, sino cómo hacer que esa configuración sea consciente.

Schön (1983) advirtió hace décadas sobre los riesgos de la racionalidad técnica en las profesiones proyectuales y propuso como alternativa la figura del practicante reflexivo. La IAG escala ese riesgo: si antes lo habitual era replicar tipologías convencionales sin cuestionarlas, ahora un estudiante puede producir miles de variantes en una tarde sin que ninguna haya sido examinada críticamente. Dunne y Raby (2013) sugirieron que el diseño

especulativo podía operar como una forma de investigación crítica. Esta investigación traslada esa intuición al terreno pedagógico y convierte la narrativa especulativa en el dispositivo que organiza la interacción del estudiante con la máquina (Figura 1).

Figura 1

Marco teórico-referencial del modelo METANOIA AI



Nota. Los cuatro cuadrantes articulan las bases conceptuales de la investigación: el tecnohumanismo como fundamento filosófico (Floridi, Sadín), la inteligencia expandida como operacionalización del acoplamiento humano-IA (Clark y Chalmers), la metanoia como objetivo pedagógico transformador (Mezirow) y el storytelling como catalizador de la transformación cognitiva (Bruner, Ricoeur, Tharp y Tharp). Las imágenes centrales, generadas mediante IA durante el taller, ilustran la convergencia entre intención humana y producción algorítmica que el marco busca problematizar.

Método

El estudio se realizó en el programa *Refugio para el cuerpo y la imaginación*, un taller de diseño avanzado a nivel de posgrado en la Universidad Michoacana de San Nicolás de Hidalgo, México. El campo disciplinar es la arquitectura, pero la lógica del modelo —usar narrativa como andamiaje crítico ante la IA— no depende de contenidos arquitectónicos específicos. Las páginas que siguen reportan el método de intervención, los resultados obtenidos con dos casos de estudio contrastantes y una discusión sobre las implicaciones para la integración crítica de la IA en educación superior.

Diseño de la investigación

Se trabajó con investigación-acción participativa (Kemmis y McTaggart, 2005). La elección no fue incidental: integrar IA generativa en el aula era territorio inexplorado para el equipo docente, y se necesitaba un diseño lo bastante flexible como para ajustar las estrategias sobre la marcha conforme se observaban las respuestas de los estudiantes. El marco conceptual incorporó además elementos del diseño discursivo de Tharp y Tharp (2019), quienes conciben los artefactos de diseño no como objetos funcionales sino como dispositivos que provocan reflexión y posicionamiento crítico. Se realizaron tres talleres experimentales entre 2023 y 2024, con 18 estudiantes de posgrado en total.

Participantes

Se convocó a estudiantes de maestría en diseño arquitectónico avanzado con al menos dos años de experiencia profesional. Los criterios de inclusión fueron familiaridad con software de modelado 3D, interés demostrado en aproximaciones especulativas al proyecto y diversidad de orígenes culturales y formativos. La diversidad de perspectivas enriqueció notablemente la identificación de sesgos algorítmicos durante la fase de curación crítica, como se describe en los resultados.

La muestra (n = 18, distribuida en tres cohortes de seis) responde al carácter intensivo de la intervención: cada taller demandó acompañamiento individualizado en las fases de iteración y curación.

Estructura del taller

El modelo pedagógico, denominado *METANOIA AI*, se implementó como taller intensivo que combinó sesiones presenciales con trabajo asincrónico. La estructura operativa distribuyó cinco días en tres bloques: uno de introducción teórico-conceptual (diseño discursivo, narrativa visual, marcos éticos de la IA), dos de exploración discursiva con IAG, y dos de materialización de las narrativas en propuestas arquitectónicas mediante modelado 3D. Se emplearon *GPT-4o* para la construcción de narrativas textuales, y *Stable Diffusion v3* junto con *Midjourney V.6* para la generación de imágenes. La coexistencia de dos generadores visuales fue deliberada: permitió a los estudiantes contrastar lo que llamamos *estéticas algorítmicas*, es decir, las tendencias visuales características de cada modelo (Figura 2).

Figura 2

Modelo pedagógico METANOIA AI: ciclo de co-creación crítica humano-IA mediante storytelling

Nota. El diagrama representa las cuatro fases del proceso —prompt narrativo, iteración dialógica, curación crítica y síntesis disciplinar— organizadas como ciclo no lineal con retroalimentación entre fases. El núcleo central señala la cognición extendida éticamente orientada como principio articulador, mientras que las tres dimensiones transversales (pensamiento crítico, conciencia ética y agencia creativa) operan simultáneamente a lo largo de todo el proceso. Las flechas de flujo (genera, selecciona, traduce, retroalimenta) indican la dirección predominante del ciclo sin excluir iteraciones regresivas entre fases. Elaboración propia a partir de la implementación del programa *Refugio para el cuerpo y la imaginación*, UMSNH 2023-2024.

Dentro de esta estructura operativa, la interacción humano-IA se articuló en cuatro fases. Primera fase: *el prompt como pregunta narrativa*. Se pidió a los estudiantes abandonar la lógica convencional de la ingeniería de *prompts* —describir una imagen deseada— y construir en su lugar el núcleo de un relato: un personaje con necesidades y contradicciones, un conflicto espacial, un escenario cultural concreto. La diferencia no es trivial. Un *prompt* descriptivo dice «casa moderna en el bosque». Un *prompt* narrativo dice: «un astrónomo retirado busca un refugio para observar las estrellas y escribir sus memorias, en un claro que perteneció a su familia, un lugar que debe ser a la vez observatorio y santuario de recuerdos, pero que también necesita acoger a sus nietos en vacaciones». El segundo obliga a la IA a negociar con una intención compleja en lugar de llenar un molde visual.

Segunda fase: *la iteración como diálogo*. Las imágenes generadas no se trataron como productos terminados sino como respuestas dentro de una conversación abierta. Cada imagen fue un interlocutor visual que el estudiante interpretó, cuestionó y ante el cual respondió reformulando su *prompt*. A lo largo del taller, los proyectos acumularon entre 67 y 87 iteraciones cada uno, documentadas en bitácoras de proceso. Lo que importó no fue la cantidad sino la evolución cualitativa de los prompts: de solicitudes genéricas a preguntas cada vez más específicas y éticamente informadas.

Tercera fase: *la curación como acto crítico*. Los estudiantes asumieron un rol curatorial: seleccionaron, recombinaron y editaron fragmentos de múltiples generaciones para componer un collage visual que condensara su narrativa. Aquí se activaron protocolos para detectar sesgos algorítmicos —la tendencia de la IA a replicar arquetipos occidentales, a preferir ciertos materiales, a idealizar relaciones entre lo humano y lo natural— y se implementó una bitácora de coautoría donde cada decisión quedó registrada según su origen: humano, algorítmico o negociado.

Cuarta fase: *la síntesis disciplinar.* El collage curado se convirtió en documento guía para el proyecto arquitectónico. Los estudiantes tradujeron atmósferas y conceptos narrativos a lenguaje técnico: plantas arquitectónicas, secciones, detalles constructivos, especificaciones materiales. Se incluyó esta fase para garantizar que la exploración especulativa no se desvinculara de las competencias disciplinares.

Instrumentos

La recolección de datos combinó técnicas cualitativas y cuantitativas. En lo cualitativo: observación participante con registro de interacciones durante las sesiones, análisis de artefactos digitales (narrativas textuales, imágenes generadas, modelos 3D, bitácoras) y entrevistas semiestructuradas sobre cómo percibían los estudiantes el papel de la IA en su proceso creativo. En lo cuantitativo: cuestionarios previos y posteriores al taller, y análisis de frecuencias extraídas de las bitácoras (iteraciones por sesión, tiempo de reflexión entre prompts, volumen de imágenes generadas). Las bitácoras resultaron la fuente más rica: al ser documentos de trabajo —no instrumentos de evaluación formal—, los estudiantes escribían con menor autocensura, lo que permitió acceder a reflexiones que difícilmente habrían emergido en un cuestionario estructurado.

Los proyectos se evaluaron en cinco dimensiones diseñadas para el estudio: discursiva (integración texto-imagen), narrativa visual (tensión visual y coherencia), expresiva (manejo de metáfora y abstracción), conceptual (generación de ideas especulativas) y de diseño (equilibrio entre lo especulativo y lo construible). Cada dimensión se calificó en escala de 1 a 5. Las rúbricas ponderaron coherencia narrativa (30%), profundidad del diálogo con IA (25%), crítica curatorial (25%) y viabilidad técnica (20%). La distribución responde a la convicción de que, en la etapa formativa, importa más cómo se piensa el proyecto que cómo se resuelve formalmente. Los resultados se contrastaron con un grupo de referencia (n = 12, estudiantes del mismo programa en semestres previos) que utilizó las mismas herramientas de IA pero sin mediación narrativa.

Resultados

Transformaciones en el aprendizaje

Los tres talleres (n = 18) arrojaron resultados convergentes en los tres ejes que estructuraron la evaluación.

Respecto a la superación del síndrome de la instantaneidad, la selección de imágenes fundada exclusivamente en criterios estéticos superficiales se redujo un 60% frente al grupo control. Los estudiantes del grupo experimental justificaron sus elecciones de diseño con argumentos conceptuales, no solo con preferencias visuales. La diferencia resultó estadísticamente significativa ($p < 0.001$, $d = 1.2$), lo que indica un efecto de gran tamaño.

En lo referente al pensamiento curatorial crítico, la transformación fue cualitativa antes que cuantitativa. Al inicio del taller, la mayoría de los estudiantes operaba la IA como se navega un catálogo: generaban imágenes y seleccionaban las más atractivas. Hacia el final, el 85% había desarrollado criterios propios de selección, combinación y reinterpretación de lo que la máquina proponía, documentados en sus bitácoras. Las mejoras en metacognición fueron significativas ($F(1,16) = 23.4$, $p < 0.001$), dato coherente con lo que Paz-Baruch et al. (2025) reportan sobre el vínculo entre andamiaje metacognitivo y pensamiento creativo.

El tercer eje fue un incremento del 40% en diversidad conceptual. Los proyectos desarrollados con mediación narrativa exhibieron mayor variedad de soluciones y mayor densidad argumental que los del grupo control. Esto sugiere que el andamiaje narrativo no solo organiza la interacción con la IA sino que amplía el espacio de búsqueda creativa al

obligar a los estudiantes a explorar ramificaciones de su relato en lugar de iterar sobre una sola imagen (Figura 3).

Figura 3

Evaluación multidimensional: promedios grupales y contraste entre casos de estudio

Escala 1-5 (1 = incipiente, 5 = sobresaliente). Talleres 2023-2024 (n = 18)

Dimensión evaluada	Promedio grupal (n = 18)	«El Susurro del Bosque»	«Un Silencio Triple»	Diferencia entre casos
Narrativa visual	**4.14**	**5**	3	+2
Conceptual	**4.14**	**5**	3	+2
Diseño	4.08	4	2	+2
Expresiva	3.86	4	3	+1
Discursiva	3.71	**5**	2	+3

La dimensión discursiva presenta simultáneamente la puntuación grupal más baja (3.71) y la mayor brecha entre casos (+3), lo cual sugiere que la integración texto-imagen en el espacio arquitectónico es el punto donde la mediación narrativa tiene mayor potencial transformador pero también donde la variabilidad individual es más pronunciada.

Elaboración propia. Rúbricas de evaluación ponderadas: coherencia narrativa 30%, profundidad del diálogo con IA 25%, crítica curatorial 25%, viabilidad técnica 20%.

Nota. Se presentan las puntuaciones en escala 1-5 para las cinco dimensiones evaluadas (narrativa visual, conceptual, diseño, expresiva y discursiva), comparando el promedio del grupo (n = 18) con los dos casos analizados en profundidad. Los círculos sólidos señalan puntuaciones sobresalientes y los delineados las más bajas. La dimensión discursiva concentra simultáneamente el promedio grupal más bajo (3.71) y la mayor brecha entre casos (+3), lo cual sugiere que la integración texto-imagen en el espacio arquitectónico es el punto de mayor potencial transformador y, al mismo tiempo, de mayor variabilidad individual. Rúbricas de evaluación ponderadas: coherencia narrativa 30%, profundidad del diálogo con IA 25%, crítica curatorial 25%, viabilidad técnica 20%.

Evaluación multidimensional

Los promedios por dimensión (escala 1-5) fueron: narrativa visual 4.14, conceptual 4.14, diseño 4.08, expresiva 3.86 y discursiva 3.71. El patrón merece atención. Las dimensiones donde el storytelling incide más directamente —la capacidad de construir tensión visual y de articular ideas especulativas— obtuvieron las puntuaciones más altas. La dimensión discursiva, que exige integrar texto e imagen en la lógica del espacio arquitectónico, resultó

la más baja: la mediación narrativa fortalece el pensamiento conceptual pero no resuelve por sí sola la dificultad de traducir ideas abstractas a propuestas espaciales.

La variabilidad entre proyectos fue considerable. *El Susurro del Bosque* y *El Refugio Reflexivo* alcanzaron 5/5 en narrativa visual; *Un Silencio Triple* obtuvo 2/5 en varias dimensiones. Esta dispersión no invalida el modelo —ninguna intervención pedagógica produce resultados uniformes—, pero sí evidencia que la efectividad de la mediación narrativa depende de la madurez conceptual previa del estudiante, su disposición a la reflexión pausada y su capacidad de verbalización.

Caso 1: «El Susurro del Bosque»

Figura 4

Lámina final del caso «El Susurro del Bosque: Serenidad Encantada», resultado de 74 iteraciones de co-creación humano-IA

Nota. El refugio orgánico se integra al entorno boscoso mediante envolventes curvas de materiales naturales. La composición refleja la evolución desde prompts descriptivos genéricos hacia una narrativa situada con personaje, conflicto ecológico y materialidad local. Puntuación: 5/5 en dimensiones narrativa visual y discursiva. Imágenes generadas con *Stable Diffusion* y *Midjourney*; concepto, curación y composición de la estudiante. Elaboración propia. con detalle cómo opera el modelo cuando funciona plenamente. La estudiante construyó una narrativa centrada en un refugio donde la arquitectura se articulaba como extensión de los procesos biológicos del bosque circundante. A lo largo de 74 iteraciones, lo que comenzó como un relato genérico sobre simbiosis entre construcción y ecosistema se transformó en un vocabulario formal propio: una gramática visual bioinspirada que la estudiante fue capaz de explicar y defender con argumentos técnicos y conceptuales.

Este proyecto permite observar con detalle cómo opera el modelo cuando funciona plenamente. La estudiante construyó una narrativa centrada en un refugio donde la arquitectura se articulaba como extensión de los procesos biológicos del bosque circundante. A lo largo de 74 iteraciones, lo que comenzó como un relato genérico sobre simbiosis entre construcción y

ecosistema se transformó en un vocabulario formal propio: una gramática visual bioinspirada que la estudiante fue capaz de explicar y defender con argumentos técnicos y conceptuales.

La bitácora documenta la evolución del diálogo con la IA. En las primeras sesiones, los *prompts* eran genéricos («casa de madera entre árboles»). Tras la intervención narrativa, incorporaron capas de significado: personajes con intenciones concretas, tensiones entre preservación ecológica y necesidad de refugio, referencias a materialidades locales. El tiempo de reflexión entre iteraciones pasó de 3 minutos en promedio en la fase inicial a 11 minutos en las fases intermedias. Simultáneamente, el número de imágenes generadas por sesión disminuyó. Menos producción, más deliberación.

En la iteración 52, la estudiante anotó que había dejado de pedir a la IA «un resultado mejor» y había empezado a pedirle «una respuesta distinta a la anterior». El cambio de verbo —de mejorar a diversificar— condensa la transición de un uso instrumental a un uso dialógico de la herramienta.

Durante la curación, identificó tres sesgos algorítmicos con ejemplos documentados: la IA generaba sistemáticamente imágenes con estética nórdica al solicitar entornos boscosos, prefería materiales industrializados sobre técnicas constructivas locales, e idealizaba la relación humano-naturaleza omitiendo tensiones reales. En lugar de descartar las imágenes sesgadas, las integró en su collage junto a correcciones narrativas, haciendo visible el proceso de negociación con la máquina. Obtuvo 5/5 en las dimensiones narrativa visual y discursiva.

Figura 5

Evolución del diálogo humano-IA a lo largo de 74 iteraciones: caso «El Susurro del Bosque».

Nota. La línea de tiempo superior muestra la transformación progresiva del *prompt* en tres fases —descriptivo (iteraciones 1-15), narrativo emergente (iteraciones 30-50) y dialógico crítico (a partir de la iteración 52)—, cada una con un cambio cualitativo en la naturaleza de la interacción con la IA. El diamante ♦ marca el punto de inflexión documentado en la iteración 52, donde el estudiante dejó de pedir «una imagen mejor» para solicitar «una respuesta distinta», evidenciando el tránsito del verbo «mejorar» al verbo «diversificar». Las dos curvas inferiores registran la relación inversamente proporcional entre tiempo de reflexión entre prompts (línea continua, ascendente de 3 a 11 minutos) e imágenes generadas por sesión (línea discontinua, descendente). La zona de cruce de ambas curvas señala el momento en que la reflexión superó a la producción como actividad dominante del proceso. Datos reconstruidos a partir de la bitácora de proceso.

Caso 2: «Un Silencio Triple»

Figura 6

Lámina final del caso «Un Silencio Triple», resultado de ~80 iteraciones de co-creación humano-IA

Nota. El refugio de ramas entrelazadas junto al agua articula tres dimensiones del silencio —corporal, natural y emocional—. La calidad visual fue notable desde las primeras iteraciones,

pero la exploración se concentró en variaciones formales sin giro narrativo profundo hasta la intervención docente en la iteración 65. Dimensión discursiva evaluada en 2/5. Imágenes generadas con *Stable Diffusion* y *Midjourney*; concepto, curación y composición del estudiante. Un patrón elocuente. Generaba hasta 45 imágenes por día —casi el doble del promedio grupal de 20-25— y dedicaba menos de 2 minutos de reflexión entre iteraciones. La cantidad funcionaba como sustituto de la profundidad: el síndrome de la instantaneidad persistía a pesar del andamiaje narrativo.

El caso ofreció, no obstante, una pista pedagógica que no se había anticipado. Cuando se implementó una intervención complementaria —sesiones de tutoría donde se le pedía narrar oralmente su proyecto antes de abrir cualquier herramienta de IA—, la coherencia narrativa mejoró de forma visible en las dos últimas sesiones. La verbalización oral pareció operar como freno cognitivo: obligaba a organizar las ideas antes de exponerlas a la aceleración algorítmica. Al término del taller, el 30% del total de participantes mostró dependencia residual de las sugerencias de la IA, lo cual confirma que el modelo necesita calibrarse según perfiles diversos y no puede aplicarse como fórmula uniforme.

Figura 7

Evolución del diálogo humano-IA a lo largo de ~80 iteraciones: caso «Un Silencio Triple»

Nota. A diferencia del caso «El Susurro del Bosque», la línea de tiempo muestra solo dos fases —descriptivo-acumulativo (dominante, ~80% del proceso) y reflexivo parcial (último 20%, tras intervención externa)—, sin transición espontánea entre ellas. Las tarjetas superiores documentan la trayectoria: un *prompt* estéticamente eficaz desde el inicio («refugio orgánico de ramas entrelazadas junto al agua») que sostuvo una producción de hasta 45 imágenes/día; una fase intermedia de variaciones formales sin desarrollo de personaje ni conflicto, donde el «silencio» permaneció como atmósfera y no como

pregunta; y la intervención docente en la iteración ~65 (▲), consistente en pedir al estudiante que narrara su proyecto sin mostrar imágenes, lo que hizo emerger la estructura del triple silencio como articulación conceptual tardía. Las dos curvas inferiores constituyen el dato central del diagrama: la línea de producción (discontinua) se mantiene alta y constante mientras la línea de reflexión (continua) permanece plana y baja, con apenas un leve repunte tras la intervención que no se consolida. Las curvas no se cruzan —a diferencia del cruce documentado en «El Susurro del Bosque»—, lo que evidencia visualmente que la alta intensidad de generación no se tradujo en profundidad dialógica. Datos reconstruidos a partir de la bitácora de proceso.

Hallazgos éticos

Uno de los resultados que no estaba en el guión fue la aparición espontánea de protocolos éticos personales. El 92% de los estudiantes desarrolló procedimientos propios para identificar y contrarrestar sesgos algorítmicos. No fue una tarea asignada. Surgió como consecuencia del proceso de curación crítica: al confrontar las imágenes generadas con sus narrativas culturalmente situadas, los sesgos se volvieron visibles sin necesidad de explicación teórica previa.

Los sesgos detectados con mayor recurrencia fueron la sobrerrepresentación de estéticas centroeuropeas y nórdicas en contextos naturales, la generación de espacios habitados por cuerpos demográficamente homogéneos y la preferencia por materiales industrializados frente a técnicas vernáculas. Algunos estudiantes inventaron lo que llamaron «contraprompts»: instrucciones narrativas diseñadas para subvertir las tendencias por defecto de la IA. La conciencia ética no fue inyectada desde fuera; el andamiaje pedagógico creó las condiciones para que emergiera desde la propia práctica de curación.

La bitácora de coautoría cumplió un papel doble. Como instrumento ético, hizo explícita la distribución de agencia en cada decisión. Como instrumento metacognitivo, obligó a preguntarse, iteración tras iteración, si una elección era propia, sugerida o negociada. Esa pregunta recurrente parece ser uno de los mecanismos centrales de la transformación que aquí se documenta. Lo que Pizaña Ramos y Martín López (2020) registraron como intensificación de la copia cuando las herramientas digitales carecen de mediación pedagógica tiene un paralelo directo en el uso de IA generativa: sin un dispositivo que obligue a explicitar las decisiones, la frontera entre lo propio y lo generado se difumina. La bitácora funciona como antídoto —no porque impida la copia, sino porque hace costoso el autoengaño.

Discusión y conclusiones

Los datos apuntan en una dirección consistente: la narrativa funciona como andamiaje cognitivo para la interacción crítica con IA generativa. La reducción del 60% en elecciones estéticas superficiales, el incremento del 40% en diversidad conceptual y el desarrollo de criterios curatoriales propios en el 85% de los participantes convergen en un patrón de reorientación en la relación del estudiante con la tecnología.

El síndrome de la instantaneidad, tal como se observó en este estudio, no es privativo de la arquitectura. Cualquier disciplina que integre IA generativa —escritura académica asistida por modelos de lenguaje, producción audiovisual con generadores de video, programación con asistentes de código— enfrenta la misma fricción entre facilidad de producción y hondura reflexiva. Lo que este estudio sugiere es que la velocidad de la IA no es dañina en sí misma; lo problemático es la ausencia de un marco que le dé intencionalidad. La narrativa cumple esa función: ralentiza estratégicamente la interacción al exigir que cada iteración responda a una pregunta con sentido, no a una búsqueda aleatoria de resultados visualmente satisfactorios.

La noción de mente extendida permite leer estos hallazgos con mayor precisión. Si la cognición se distribuye entre cerebro, cuerpo y entorno (Clark y Chalmers, 1998; Paul, 2021), entonces el sistema estudiante-IA es, en sentido técnico, un sistema cognitivo. Pero no cualquier acoplamiento produce cognición de calidad. Lo que distingue al acoplamiento productivo del uso meramente instrumental es la presencia de un mediador crítico. Aquí, la narrativa fue la estructura que convirtió una interacción reactiva en diálogo reflexivo. En términos de la postfenomenología de la mediación tecnológica (Verbeek, 2015), el storytelling estableció una relación hermenéutica con la IA: la máquina se volvió algo que se interpreta, no algo que se obedece.

Conviene, sin embargo, resistir la tentación de romantizar el hallazgo. El caso de «Un Silencio Triple» recuerda que la narrativa no es varita mágica. Su eficacia depende de condiciones que este estudio apenas comienza a mapear: la disposición del estudiante a la reflexión pausada, su capacidad para verbalizar intenciones antes de ejecutarlas, quizá también su relación previa con la lectura y la escritura como prácticas intelectuales. Que la verbalización oral funcionara como intervención de rescate sugiere que existen múltiples vías de acceso al pensamiento narrativo, y que la escritura no es la única. Este hallazgo

dialoga con lo que Paz-Baruch et al. (2025) reportan sobre la necesidad de adaptar las intervenciones metacognitivas a los perfiles individuales.

Las limitaciones del estudio son claras. La muestra es pequeña y corresponde a un posgrado específico. El diseño de investigación-acción implica que quien investiga es también quien facilita, con el sesgo que eso conlleva. Y las herramientas de IA generativa evolucionan a un ritmo que vuelve perecederos los detalles técnicos. Lo que no perece es la lógica subyacente: usar la narrativa como mediador crítico para la interacción con cualquier IA generativa. Esa lógica podría adaptarse a diseño industrial, artes visuales, comunicación o formación docente. Estudios futuros deberían probar esa transferibilidad con diseños cuasi-experimentales, muestras más amplias y seguimiento longitudinal del impacto en la práctica profesional.

Las implicaciones van más allá de lo disciplinar. La educación superior enfrenta globalmente la necesidad de integrar IA sin sacrificar pensamiento crítico, y este estudio sugiere que la clave no reside en regular el acceso a la tecnología ni en enseñar sus funcionalidades, sino en diseñar andamiajes cognitivos que sostengan la agencia del estudiante sobre el proceso creativo. En cualquier campo donde se generen artefactos con IA —textos, imágenes, código, diseños—, la pregunta pedagógica de fondo sigue siendo la misma: ¿cómo evitar que el estudiante se convierta en selector de opciones algorítmicas?

El concepto de curación crítica que emerge de esta investigación —seleccionar, combinar, interpretar y responsabilizarse éticamente de los productos de la co-creación con IA— podría constituir una competencia transversal para la educación superior contemporánea, análoga a lo que la alfabetización informacional representó para generaciones anteriores.

Para cerrar, regresamos a la bitácora de aquella estudiante que dijo dialogar con su sombra. Lo que nos parece significativo no es la elocuencia de la frase sino el momento en que apareció: después de sesenta iteraciones, cuando la acumulación de experiencia con la máquina había creado las condiciones para la reflexión. Nadie le explicó qué era un sesgo algorítmico. Lo descubrió narrando. Los datos de este estudio sugieren que eso no fue accidental: la narrativa, cuando se estructura como dispositivo pedagógico, produce transformaciones medibles en la relación con la IA, en el pensamiento crítico y en la sensibilidad ética. Quizá lo más valioso que este trabajo pone de manifiesto sea algo elemental: que enseñar con IA exige, ante todo, enseñar a formular las preguntas que las máquinas no saben hacer.

Figura 8

Síntesis de la discusión: del síndrome de la instantaneidad a la curación crítica como competencia transversal

Nota. El diagrama presenta la lógica argumentativa del estudio en tres columnas —diagnóstico, intervención y resultado— articuladas por la pregunta central de la investigación. La columna de intervención integra los referentes teóricos (Verbeek, Clark y Chalmers) como mecanismos explicativos, no como catálogo. La franja inferior señala los campos disciplinares donde la lógica del andamiaje narrativo resulta transferible.

Referencias

Akolekar, S., Banerjee, A., Kumar, R. y Patel, M. (2025). Evaluating large language models in engineering education: A comprehensive assessment. *Nature Scientific Reports*, *15*(1), 1-14.

Bruner, J. (2003). *La fábrica de historias: Derecho, literatura, vida*. Fondo de Cultura Económica.

Chen, M., Liu, X., Wang, S., & Torres, R. (2024). Longitudinal impact of AI tools on architectural design education. *International Journal of Educational Technology in Higher Education*, *21*(3), 78-95.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7-19.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Dunne, A., & Raby, F. (2013). *Speculative everything: Design, fiction, and social dreaming*. MIT Press.

Kemmis, S., & McTaggart, R. (2005). Participatory action research: Communicative action and the public sphere. En N. Denzin e Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed.) (pp. 559-603). SAGE.

Leach, N. (2022). *Architecture in the age of artificial intelligence: An introduction to AI for architects*. Bloomsbury Visual Arts.

Lobato-Valdespino, J. C., & Flores Romero, J. H. (2024). Más allá del estado estable: diseño discursivo como práctica reflexiva asistida por IA. En *Actas de las Jornadas sobre Innovación Docente en Arquitectura* (JIDA'23). Iniciativa Digital Politècnica UPC.

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, *33*(4), 659-684.

Paul, A. M. (2021). *The extended mind: The power of thinking outside the brain*. Houghton Mifflin Harcourt.

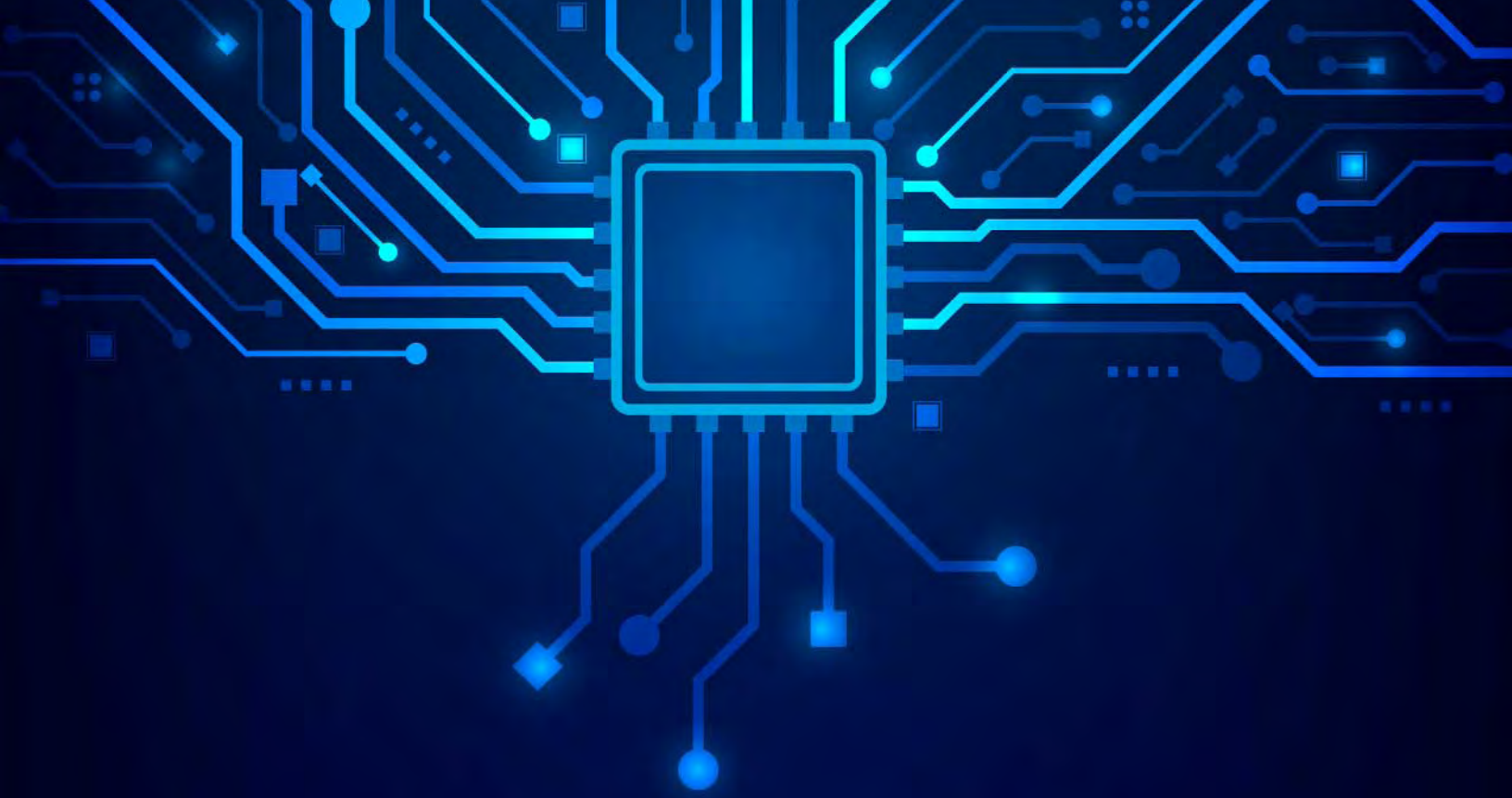
Paz-Baruch, N., Grovas, M. y Mevarech, Z. (2025). The effects of meta-creative pedagogy on elementary school students' creative thinking. *Metacognition and Learning*, *20*(1), 45-68.

Pizaña Ramos, L. E., & Martín López, L. (2020). *Sobre la copia y el plagio dentro de los talleres de proyectos de arquitectura*. ASINEA.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Tharp, B. M., & Tharp, S. M. (2019). *Discursive design: Critical, speculative, and alternative things*. MIT Press.

Verbeek, P. P. (2015). Beyond interaction: A short introduction to mediation theory. *Interactions*, *22*(3), 26-31.



20.

Aplicación de la inteligencia artificial en la gestión administrativa de las micro, pequeñas y medianas empresas del sector de la construcción

Elizabeth Vanessa Teniente Gasca
Universidad Latina de México
ORCID: 0009-0009-7090-4586

Octavio Reyes López
Universidad del Centro del Bajío, México
ORCID: 0000-0003-3960-1228

Christian Paulina Mendoza Torres
Tecnológico Nacional de México / Instituto Tecnológico de Celaya
ORCID: 0000-0002-2998-714X

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.20

20.

Aplicación de la inteligencia artificial en la gestión administrativa de las micro, pequeñas y medianas empresas del sector de la construcción

Introducción

La presente investigación se desarrolló con el objetivo de estudiar la aplicación de las herramientas de inteligencia artificial (IA), para integrar los factores de la gestión administrativa de las micro, pequeñas y medianas empresas (MiPyME), del sector de la construcción en la región del Bajío mexicano, a través de un acercamiento hermenéutico y fenomenológico apoyado en la Estructura Metodológica Fundamentada para Investigaciones Sociales (EMFIS v-5).

La teoría clásica de la administración, propuesta por Henry Fayol desde 1916 (Arguello et al., 2020), junto con las aportaciones de George Terry (1962), establecen que la gestión administrativa de una organización debe desarrollarse a través de cuatro etapas básicas para dicho proceso administrativo: Planeación, Organización, Ejecución y Control.

De manera complementaria, la gestión administrativa de una empresa clásica se desarrolla mediante cuatro Áreas Funcionales: Contabilidad y Finanzas; Mercadotecnia y Ventas; Gestión de Operaciones o Producción; así como Manejo de Personal (López, 2010).

En México, las MiPyME se enfrentan a la problemática y los desafíos de ajustar su gestión administrativa a los modelos teóricos tradicionales, los cuales se apoyan predominantemente en las Etapas del Proceso Administrativo, en conjunción con las Áreas Funcionales Clásicas de una organización, lo cual representan esquemas de rigidez en la estructura, altos costos de operación y lentitud en la toma de decisiones cotidiana, reduciendo así su nivel de competitividad (Armenteros et al., 2024; Salinas, 2024).

Habitualmente, las MiPyME no cuentan con mucho personal para trabajar en las Áreas Funcionales de la gestión empresarial, que son la Contabilidad y Finanzas, la Mercadotecnia y Ventas, la Gestión de Operaciones, así como el Manejo de Personal (López, 2010), razón por la cual éstas, son desatendidas en un intento por mejorar la rentabilidad de subsistencia. Esta desintegración limita una planeación que acceda a la sostenibilidad y competitividad a largo plazo (Zabaleta et al., 2024), por lo que la ejecución y el control suelen ser propensos a errores humanos, limitando así la capacidad de respuesta ante los entornos cambiantes (Salinas, 2024; Soledispa et al., 2022).

Por lo anterior, se considera viable y factible la integración de las Etapas del Proceso Administrativo con las Áreas Funcionales de la Organización, que apoyadas en la IA así también con la inteligencia artificial generativa (IAG), misma integración que se presenta en este reporte como un eje disruptivo capaz de transformar radicalmente a la gestión administrativa (Teniente, 2026); ya que la IA actúa como consultor estratégico que asiste al ser humano con el diseño de organigramas dinámicos, la redacción técnica y optimización de problemas y operaciones complejas (Ellingrud et al., 2023).

La importancia de implementar la IA en las MiPyME radica en mejorar la calidad, productividad, competitividad y sustentabilidad y sostenibilidad de las empresas, a través de automatizar los procesos de conocimiento y fundamentar los planes de negocio hacia los datos actuales (McKinsey & Company, 2023). No obstante, las organizaciones deben asegurarse de trabajar con criterios de gobernanza ética y evaluar la configuración de resultados para evitar los sesgos algorítmicos (Diestra et al., 2021; Mejía et al., 2025).

Método

La investigación se realizó con un enfoque cualitativo desde una perspectiva hermenéutica y fenomenológica con el propósito de integrar la opinión, experiencias y vivencias de los entrevistados dentro de su contexto organizacional, tal como lo señalan Hernández-Sampieri y Mendoza-Torres (2018). Así también, se utilizó el diseño de la Teoría Fundamentada en unión con la Estructura Metodológica Fundamentada para la Investigaciones Sociales (EMFIS, v-5) de acuerdo con Reyes y Hernández-Moncada (2019).

El proceso EMFIS ayudó a generar las categorías emergentes de estudio a través de interacción de la primera dimensión de la etapa del proceso administrativo y la segunda dimensión que son las áreas funcionales de la organización. La Estructura Metodológica

Fundamentada para Investigaciones Sociales (EMFIS), busca unir el objetivo de la investigación con la metodología para validar y asegurar la fiabilidad de los resultados (Reyes y Hernández-Moncada, 2019). Para el desarrollo del esquema EMFIS, se dio seguimiento a la siguiente estructura investigativa: a)Enfoque de la Investigación, b)Perspectiva del Estudio, c)Tipo de Investigación, d)Mapa Conceptual "a-priori", e)Diseño Estructural, f)Andamiaje Metodológico, g)Postura del Análisis, h)Coreografía de Datos, i)Población Bajo Estudio, j) Tipo y Tamaño del Muestreo, k)Instrumento de Indagación, l)Inmersión Inicial, m)Unidad de Análisis, n)Triangulación de la Información, o)Fiabilidad, p)Consistencia, q)Validación, r)Análisis del Material, s)Criterios de Codificación, t)Codificación Resultante, u)Cadena de Significación, v)Mapa Conceptual *a-posteriori*, w)Elementos Triuno, x)Matriz de Congruencia, y)Tabla de Coherencia, z)Categorías "a-posteriori" de los Factores Bajo Estudio.

El arranque de la investigación se genera desde la perspectiva de la Teoría Clásica de Henry Fayol (1916, citado por Arguello et al., 2020), el cual define la naturaleza de la Administración y George Terry (1962) quien define el Proceso Administrativo, por lo cual la primera dimensión de estudio fueron las etapas del Proceso Administrativo: Planeación, Organización, Ejecución y Control. Por lo cual las Áreas Funcionales de la Gestión Empresarial, representan la segunda dimensión de estudio: Contabilidad y Finanzas, Mercadotecnia y Ventas, Gestión de Operaciones, Manejo de Personal (López, 2010).

En la Figura 1 se presenta la interacción emergente de las Etapas del Proceso Administrativo con las Áreas Funcionales en la Gestión Empresarial representado de la siguiente manera:

La orbita interior representa la Etapas del Proceso Administrativo (E).

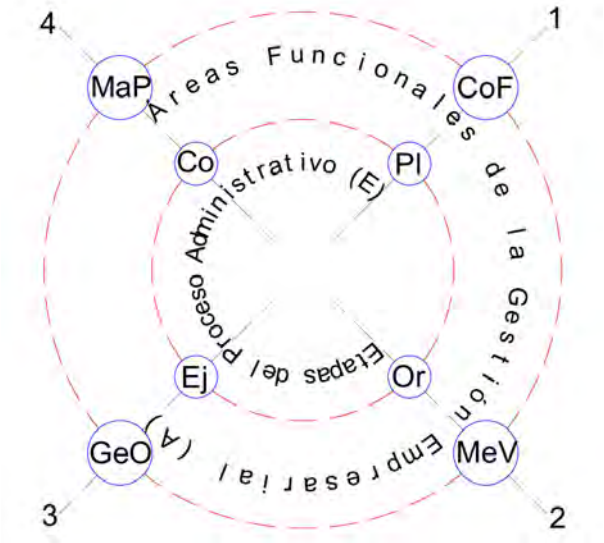
La orbita exterior simboliza las Áreas Funcionales de la Organización (A).

- El Eje 1 se encuentran la Planeación (Pl) y Contabilidad y Finanzas (CoF).
- En el Eje 2 se plasma la Organización (Or) y Mercadotecnia y Ventas (MeV).
- En el Eje 3 manifiesta la Ejecución (Ej) y Gestión de Operaciones (GeO).
- El Eje 4 expresa el Control (Co) y Manejo de Personal (MaP).

Entre las orbitas y ejes se crean interacciones entre las Etapas del Proceso Administrativo y las Áreas Funcionales de las organizaciones creando una convergencia entre ellas, en la cual el trabajo conjunto mejora la Gestión Empresarial para las MiPyME.

Figura 1

Interacción emergente de las Etapas del Proceso Administrativo con las Áreas Funcionales de la Organización



Nota. Adaptado de Teniente (2026, p. 52).

Mediante un conjunto de entrevistas apoyadas en un enfoque de análisis cualitativo con perspectiva hermenéutica y fenomenológica (Hernández-Sampieri & Mendoza-Torres, 2018), se desarrolló la interpretación sobre la opinión de los participantes, en dónde emergieron diversos factores de cómo la tecnología puede apoyarles.

La fusión estratégica lograda al integrar las Etapas del Proceso Administrativo con las Áreas Funcionales de la Organización, se favoreció la comprensión visual por parte de los entrevistados, para proceder a interrogarles sobre las alternativas de aplicación de la IA en la Gestión Administrativa (Teniente, 2026).

En la Tabla 1 se expresa la estructura de las Etapas del Proceso Administrativo y las Áreas Funcionales de la organización teniendo como punto de partida la figura 1, consolidando el constructo teórico de la fusión de la Inteligencia Artificial con la teoría clásica, representando una matriz en donde se encuentran las dos dimensiones anteriormente definidas.

Tabla 1

Matriz de interacciones de las dimensiones de la Teoría Clásica para aplicar la IA en la Gestión Administrativa

Etapas del Proceso Administrativo (E)	Áreas Funcionales de la Gestión Empresarial (A)			
	Contabilidad y Finanzas (CoF)	Mercadotecnia y Ventas (MeV)	Gestión de Operaciones (GeO)	Manejo de Personal (MaP)
Planeación (Pl)	EA11	EA12	EA14	EA14
Organización (Or)	EA21	EA22	EA24	EA24
Ejecución (Ej)	EA31	EA32	EA34	EA34
Control (Cr)	EA41	EA42	EA44	EA44

Nota. Adaptado de Teniente (2026, p. 51).

De la codificación compartida, se procedió a la triangulación de la información para asegurar la fiabilidad, consistencia y validación de las categorías emergentes; en dónde la cadena de significación proporcionó diversos factores para su estudio.

Primeramente, es mester afirmar qué, la planeación es fundamental para establecer objetivos, diagnósticos, presupuestos y metas de las empresas, requiriendo adaptarse a los cambios sin afectar la dirección inicial (Arguello et al., 2020), al integrar la IA en la planeación de una empresa se logra una toma de decisiones en tiempo real, por lo que promueve la innovación y su capacidad de respuesta ante los entornos variables de una microempresa (Salinas, 2024), por lo que la interrelación de la planeación administrativa con la IA, expresa un progreso para las organizaciones hacia un futuro, la IA ayuda a gestionar los recurso de las MiPyME.

Con el apoyo de la IAG, actúa como consultor estratégico en la etapa de planeación, ya que asiste en el diseño de organigramas dinámicos, redacción técnica y optimización de operaciones (Ellingrud et al., 2023). Al implementar la IAG permite a las MiPyME automatizar procesos del conocimiento y generar contenido variado con el objetivo de fundamentar el plan de negocio y los datos actuales (Mckinsey y Company, 2023). Es importante que estos procesos cuenten con supervisión humana, ya que la IAG es una herramienta que permite que aumente la productividad y promueva la creatividad estratégica, el humano asegura una gobernanza ética y confirmación de resultados (Diestra et al., 2021).

La planeación en conjunto con la IA favorece la sostenibilidad y competitividad a largo plazo de las MiPyME tradicionales (Zabaleta et. al., 2024). Estos algoritmos de aprendizaje

generan datos predictivos en demanda y rotación de inventario, así como ayudan a prevenir problemas y elevar la rentabilidad (Montenegro, 2024; Soledispa et al., 2022), por lo cual la planeación se convierte en un proceso dinámico obteniendo una mejora continua, donde la intervención de la tecnología en conjunto con el ser humano converge entre sí para alcanzar las metas de una organización (Mejía et al., 2025; Soto, 2021).

La organización desde el punto de vista de la teoría clásica su función principal es disponer y coordinar los recursos humanos, materiales y financieros de la empresa (Arguello et al*.,* 2020), por lo cual es un elemento esencial para la división del trabajo, asignación de responsabilidades y el fortalecimiento de los canales de comunicación en una empresa (Mendoza y Moreira 2021), la correlación entre la organización y la IA es un cambio de paradigmas para la evolución estructural de la gestión administrativa ya que permite tener una mayor disposición y coordinación de recursos más dinámico y eficiente.

Mediante la IAG los diseñadores pueden diseñar matrices funcionales y procedimientos técnicos que puedan ser utilizados en tiempo real, asegurando la coordinación de esfuerzo para obtener una ventajea competitiva sostenible (Sarmiento et al., 2024; Muñoz, 2024).

La ejecución en una empresa permite que los procedimientos sean acciones enfocadas para generar eficacia y eficiencia en la empresa la cual se basa en metas institucionales, se adoptan estrategias de acuerdo con las necesidades de organización e identificación de oportunidades en las organizaciones permitiendo determinar medidas correctivas para llegar a una solución (Soto, 2021). Al tener una acción con la IA la ejecución el proceso operativo cambia para ser dinámico en el cual se identifican las deficiencias y se determinan las medidas correctivas en su solución (Ellingrud et al., 2023).

El control es una función crítica dentro de las empresas ya que asegura que operen de una manera eficiente con el seguimiento de actividades y corrección de remociones (Soledispa et al*.,* 2022), con ayuda de la IA el control se ve potenciado ya que las inspecciones y auditorias se pueden realizar de manera sistemática y oportuna en los recursos materiales y financieros (Arguello et al., 2020). En el ámbito financiero la IA permite minimizar errores humanos con la finalidad de tener una precisión contable, las cuales facilitan las conciliaciones bancarias de manera casi instantánea (Sánchez et al., 2024).

Al integrar la IAG con la ejecución mejora el diseño de productos, optimiza operaciones complejas y genera resultados de acuerdo con las necesidades detectadas, por lo que

la auditoría periódica ayuda a detectar fallos algorítmicos y garantizar que las decisiones automatizadas se mantengan dentro de un marco ético y profesional (Diestra et al., 2021; Aparicio et al., 2023).

La contabilidad mide el rendimiento de una empresa, por lo que la IA ayuda a optimizar registros económicos y planificación financiera basados en datos históricos de datos de la organización (López et al., 2024). Los sistemas inteligentes ayudan a minimizar los errores humanos y mejorar la precisión contable de las MiPyME (Sánchez et al., 2024). Entre los beneficios más tangibles generados con la IAG es la capacidad de realizar concordancia de las cuentas bancarias de manera automática, lo que permite acelerar el cierre mensual y anual, permitiendo acelerar la obtención de fornos y análisis de rentabilidad en el área financiera (Sánchez et al., 2024; Diestra et al., 2021). Es importante denotar que la IA conlleva a desafíos importantes relacionados con la ciberseguridad y la privacidad de datos financieros sensibles (Aparicio et al., 2023), por lo que es importante que la IA se utilice como complemento para aumentar la productividad del personal, requiriendo la supervisión de expertos para interpretar los resultados complejos y garantizar la rendición de cuentas en la gestión financiera (McKinsey y Company, 2023; Sánchez et al., 2024).

La mercadotecnia y ventas son parte fundamental de una empresa por encargarse del posicionamiento de la marca creando incentivos que promuevan las ventas para poder consolidar a las organizaciones (Cardona et al., 2022). En conjunto con la IA, la mercadotecnia ayuda a integrar y optimizar procesos comerciales, manejando grandes volúmenes de información ya que identifica patrones de consumo, impulsando la adaptabilidad ante los cambios del mercado (Salinas, 2024), y permitiendo una segmentación, así como una personalización de campañas más efectivas para llegar a nuevos clientes, mejorando el conocimiento de la marca (Armenteros et al., 2024; Ellingrud et al., 2023).

La IAG facilita la creación de contenido publicitario con elementos diferenciadores que al funcionar los chatbots o agentes conversacionales proporcionan un mejor servicio al cliente rápido y eficiente, aumentando la competitividad empresarial (Ellingrud et al., 2023; Montenegro, 2024). Por lo que al combinar los algoritmos de aprendizaje en las organizaciones permite optimizar la cobertura geográfica, reduciendo costos para la entrada de nuevos mercados, las cuales superan las barreras tradicionales como son el desconocimiento del consumidor actual (Armenteros et al.,2024; Flores et al., 2020).

De acuerdo con Rodríguez et al. (2023) la gestión de operaciones es un conjunto de procesos que intervienen dentro la planificación, la organización, la implementación y control de actividades que tienen relación con la producción de bienes y servicios para alcanzar niveles recomendables de competitividad. La analítica 4.0 de la IA analiza los procesos productivos integrando el aprendizaje automático para la toma de decisiones estratégicas y mejorar la autonomía en la ejecución de los datos (Ortega, 2024). El conjunto de las tecnologías disruptivas beneficia la automatización de tareas repetitivas y mecánicas que de manera tradicional era importante la intervención humana, por lo que elevan la productividad del personal operativo (Tyler et al., 2023; Ellingrud et al., 2023).

Debido a que existen limitaciones y posibles sesgos en los sistemas automatizados es importante trabajar en sinergia entre la tecnología y el juicio humano, realizando auditorías periódicas para garantizar la transparencia, la seguridad de la información y la ética dentro de la estructura operativa (Diestra et al., 2021; Mejía et al., 2025).

Tradicionalmente, el manejo de personal coordina esfuerzos comunes como son el liderazgo y la motivación para crear un sistema saludable dentro de la organización. La IA ayuda a alinear los objetivos individuales del personal con las metas institucionales (Sarmiento et al., 2024). Al automatizar tareas administrativas rutinarias y procesar grandes volúmenes de datos la IA identifica las cualidades del personal operativo, permitiendo mejorar el proceso de reclutamiento y la asignación de actividades, así como la programación de turnos (Espinoza, 2024; Meza y Flores, 2024).

El desarrollo de un nuevo talento humano especializado será clave para transformar los roles tradicionales por lo que la IAG da un valor estratégico al intervenir en el reentrenamiento y capacitación continua de los trabajadores hacia la nueva era digital, es así qué, la alfabetización digital permite la reducción de la brecha tecnológica, asegurando la supervivencia de las MiPyME en entornos económicos inciertos (Ellingrud et al., 2023; Ortega, 2024).

La fusión de la IAG con la gestión administrativa representa una cambio de perspectiva para las MiPyME, por tratarse de otro paradigma con un eje disruptivo para las MiPyME tradicionales, ya que con la integración de la tecnología ayuda a optimizar procesos operativos como lo son control de inventarios, solicitud de materiales, distribución de producto, mejor atención al cliente, etc., por lo tanto, mejora la toma de decisiones en tiempo real, adaptando al empresario a la transformación de la innovación empresarial.

Del método empleado emergieron las categorías *a-posteriori* de los factores bajo estudio y después de triangular la información se aseguró de esta manera, que los datos son fiables, consistentes y válidos para el propósito buscado.

Resultados

Una vez analizados los conceptos de las Etapas del Proceso Administrativo con las Áreas Funcionales de la Organización integradas con la IA y la IAG, se visualizó una convergencia interesante entre todos los conceptos estudiados, ya que la interacción proporcionada de la Planeación, Organización, Ejecución, Control, en conjunto con la Contabilidad y Finanzas, la Mercadotecnia y Ventas, la Gestión de Operaciones y el Manejo de Personal, hacen que la empresa fortalezca su operatividad llevándolos a una mejor y acertada toma de decisiones, consolidando el constructo teórico representado en la Tabla 1, que al interactuar en conjunto con la IA hacen que la toma de decisiones se convierta en una resolución casi instantánea sin perder de vista la supervisión humana para evitar sesgos éticos y algorítmicos ante lo que se plantea con la IAG, tal como se muestra en la Figura 2, la cual es inspirada en la proporción de Fibonacci, identificada cómo una contribución significativa a la numeración árabe con base decimal.

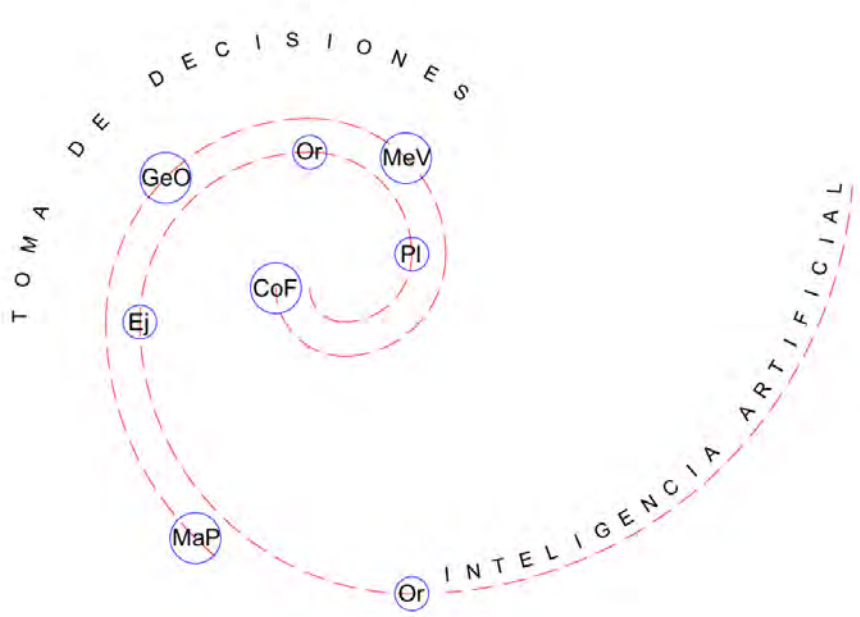
Cabe decir qué, a esta sucesión matemática también se le conoce como proporción áurea o espiral de oro (García y Martínez, 2021). Derivado del análisis previo, se integra a manera de diagrama un modelo de Gestión Administrativa que puede ser aplicado en las MiPyME para mejorar su rentabilidad y asegurar su sobrevivencia con los recursos disponibles.

El modelo visualiza una convergencia armónica de los conceptos de la Teoría Clásica, reconociendo que las MiPyME no crecen de manera lineal, si no que atraviesan por una serie de sucesos, la espiral inspira la optimización de procesos que es impulsada por la inteligencia artificial simbolizando un flujo dinámico para mejorar la toma de decisiones dentro de las MiPyME.

La geometría del modelo trabaja en conjunto con las dos dimensiones y categorías emergentes que bajo supervisión humana consolidan las operaciones de la empresa, funcionando como un marco metodológico visual, que sirve de guía hacia la transformación de las MiPyME tradicionales, permitiendo adaptarse a la innovación empresarial y mejorar su competitividad.

Figura 2

Interacción Emergente de las Etapas del Proceso Administrativo con las Áreas Funcionales de la Organización con la IAG



Nota. Adaptado de la sucesión de Fibonacci (García & Martínez, 2021).

Se aporta una Planeación dinámica mediante la integración del a IAG, para que la planeación deje de ser estática para transformarse en un proceso de mejora continua, fortaleciendo la toma de decisiones basados en datos reales y predictivos en la demanda y rotación de inventarios.

La IA identifica deficiencias operativas de manera inmediata, permite una Organización y Ejecución más eficientes mediante la coordinación y creación de matrices funcionales que aseguran una ventaja competitiva y sostenible para el empresario. Por lo cual, el control se potencializa con la capacidad del uso de la IA mediante inspecciones sistemáticas y conciliaciones bancarias automáticas, permitiendo minimizar el error humano y agilizar el análisis de la rentabilidad.

Los beneficios que aporta la fusión de la Teoría Clásica con la IA, en cuestión de finanzas ayuda a la reducción de tiempos en cierres mensuales y incrementa la precisión contable. El uso de *chatbots* junto con la analítica establece patrones de consumo y personaliza campañas, ayudando a sobrepasar las barreras en la fata de conocimiento del mercado. Con la automatización de tareas repetitivas se optimiza el reclutamiento del personal mejorando a la toma de decisiones en las operaciones y el personal.

La autonomía de los sistemas de IA no funciona de manera independiente, es importante tener una supervisión humana constante como requisito indispensable en las MiPyME, ya que el juicio humano garantiza la rendición de cuentas, la transparencia, así como la mitigación de sesgos algoritmos, obligando a la organización en apostar a la alfabetización digital como supervivencia y sostenibilidad organizacional.

Discusión y conclusiones

Ante la llegada de la IA, las MiPyME requieren un cambio de paradigmas en la Gestión Administrativa, por lo que se propone fusionar la Teoría Clásica de la Administración con la IA, siendo una necesidad de supervivencia para las organizaciones que operan de manera tradicional, partiendo de dos pilares importantes del ámbito de la Administración como lo son Fayol (1916, citado por Arguello et al., 2020) y Terry (1962) integrados con la IA adquieren dimensiones que trabajan en conjunto con la nueva era digital, integrando algoritmos de aprendizaje y modelos generativos.

De esta forma se cumple con el objetivo general propuesto para la presente investigación, al estudiar la aplicación de las herramientas de IAG (*gadgets*), para integrar los factores de la Gestión Administrativa, correspondiente a las Etapas del Proceso con las Áreas Funcionales de las MiPyME, a través de un acercamiento de tipo hermeneútico y fenomenológico apoyado en el esquema EMFIS, v-5, a fin de robustecer a las organizaciones actuales mediante sus interacciones proporcionales, guiando a las empresas a una mejor toma de decisiones casi instantánea y adaptada a la transformación digital.

Entre los principales hallazgos encontrados se puede recuperar que la IA actúa como consultor estratégico en las MiPyME ya que cambia la mirada hacia la planeación a través de una perspectiva dinámica y predictiva abierta a las transformaciones de las tecnologías disruptivas. Arguello et al. (2020) mencionan la necesidad de adaptar la planeación hacia las variaciones del entorno, demostrando que los resultados manifiestan que la integración de la IAG en las organizaciones generan una toma de decisiones casi instantánea que se fundamentan en el uso de datos en tiempo real, elevando la competitividad y rentabilidad de las organizaciones. Desde el punto de vista financiero Sánchez et al. (2024) encuentra a la automatización de tareas como un soporte en las conciliaciones bancarias y el análisis de rentabilidad, con el fin de reducir el error humano, validando la precisión contable.

Es imperativo destacar que la tecnología no desplaza al juicio humano, autores como Diestra et al. (2021) y Mejía et al. (2025) enfatizan que, aunque la IA optimiza la ejecución y el control, la supervisión humana será indispensable y fundamental para la gobernanza ética y la corrección de errores existentes en los sesgos algorítmicos. Por lo que el uso equilibrado entre la automatización y la ética profesional permite que la innovación sea sostenible a largo plazo.

En cuanto a las áreas de mercadotecnia y manejo del personal, el estudio expresa que la IA hace más fácil la personalización y la gestión del talento humano referidos a la eficiencia del personal, así como la capacidad de procesar grandes volúmenes de datos, los cuales permiten reconocer patrones de consumo y optimizar el reclutamiento, entre otras cosas, por medio de la transformación de roles tradicionales con la alfabetización digital del personal.

Generando un aporte hacia la integración del marco teórico se generó una integración de la Teoría Clásica en conjunto con la IA desarrollando un modelo de Gestión Administrativa, inspirada en la proporción de Fibonacci propuesta por Teniente (2026), fundamentada en la proporción áurea, conceptos encontrados en la publicación de García y Martínez (2021). El modelo simboliza la convergencia armónica ente las Etapas del Proceso Administrativo con las Áreas Funcionales de la empresa mediante el uso de la IA y la IAG. Generando un constructo teórico que proporciona a las MiPyME una ruta clara hacia la transición de la transformación digital, asegurando que la estructura operativa sea eficaz y eficiente, siendo resiliente ante la incertidumbre que se enfrentan las MiPyME en la actual economía global.

Referencias

Aparicio, O. Y., Ostos, O. L., & Von, O. F. (2023). Competencia digital y desarrollo humano en la era de la Inteligencia Artificial, *Hallazgos*, *20*(40), 217 – 235. https://doi.org/10.15332/2422409X

Arguello, A. M., Llumiguano, M. E., Gavilánez, C. V., & Torres, L. H. (2020). *Administración de empresas. Elementos Básicos*. Pons Publishing House.

Armenteros, N. R., Jiménez, M., & Romero, J. A. (2024). Relación entre herramientas tecnológicas, características de la empresa e ingresos en las MiPyMEs mexicanas: análisis de asociación por grupos (clústers). *Información tecnológica*, *35*(2), 37-48. https://dx.doi.org/10.4067/s0718-07642024000200037

García, R., & Martínez, L. (2021). El número de oro: Aplicaciones contemporáneas en la morfología biológica y el diseño. *Revista de Matemáticas y Naturaleza*, *12*(2), 38-52.

Cardona, D. A., Morelos, J., & Caraballo-Hernández, K. (2022). Equipos de ventas y su incidencia en el desempeño de las organizaciones: una revisión sistemática de la literatura. *Revista de Investigación, Desarrollo e Innovación*, *12*(2), 185-196. https://doi.org/19053/20278306.v12.n2.2022.15259

Diestra, N. M., Córdova, A. J., Caruajulca, C. P., Esquivel, D. L., & Nina, S. A. (2021). La inteligencia artificial y la toma de decisiones gerenciales. *Revista de Investigación Valor Agregado*, *8*(1), 52-69. https://doi.org/10.17162/riva.v8i1.1631

Ellingrud, K., Sanghvi, S., Dandona, G., Sanders, N., & Zou, W. (2023). *The state of organizations 2023: Ten shifts that are transforming organizations*. McKinsey & Company. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-state-of-organizations-2023

Espinoza, D. L. (2024). La Inteligencia Artificial como Apoyo en los Procesos de la Administración Empresarial. *Código Científico Revista de Investigación*, *8*(6), 334–364. https://doi.org/10.37811/cl_rcm.v8i6.15019

Flores, F., Méndez, J. J., & Bermúdez, C. P. (2020). Estrategias digitales para las pymes. *Revista Innova ITFIP*, *6*(1), 29–49. https://doi.org/10.54198/innova06.02

Hernández-Sampieri, R., & Mendoza-Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Interamericana Editores.

López, J. E., Gavilanes, E. A., Haro, A. F., Viteri, J. S., & Lema, L. G., (2024). *Entre Números y Teorías: La Intersección de la contabilidad con la Filosofía, la economía y las Ciencias*. Know Press. https://doi.org/10.70180/978-978-9942-7273-0-5

López, M. (2010). F*undamentos de economía, empresa, derecho, administración y metodología de la investigación aplicada a la RSC*. NETBIBLO, S. L.

McKinsey y Company, (2023). *Technology Trends Outlook 2023*. McKinsey & Company. https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/technology%20trends%20outlook%202023/technology-trends-outlook-2023.pdf

Mejía, S. E., Nava, J. E., & Cedeño, R. J., (2025). La participación de la inteligencia artificial en la toma de decisiones gerenciales. *Revista InveCom*, *5*(4). https://doi.org/10.5281/zenodo.14816449

Mendoza V. M., & Moreira, J. S. (2021). Procesos de Gestión Administrativa, un recorrido desde su origen. *FIPCAEC*, *6*(3), 608–620. https://doi.org/10.23857/fipcaec.v6i3.414

Meza, F., & Flores, O. A. (2024). La gestión administrativa y su influencia sobre la satisfacción laboral en empleados financieros en organizaciones sin fines de lucro. *Unaciencia Revista de Estudios e Investigaciones*, *17*(32), 4-21. https://doi.org/10.35997/unaciencia.v17i32.752

Montenegro, A. A., (2024). *Inteligencia artificial como herramienta de gestión en la mejora de procesos en las MiPyMEs: caso ferretería Angie*. La Libertad. [Tesis Licenciatura] UPSE, Matriz. Facultad de Ciencias Administrativas. https://repositorio.upse.edu.ec/handle/46000/10891

Muñoz, J. M. (2024). IA generativa desafíos para la propiedad intelectual. *Revista de Derecho UNED*, 33, 18-66. https://doi.org/10.5944/rduned.33.2024.41924

Ortega, F. (2024). BPM y analítica de datos en la gestión administrativa para mejorar los procesos empresariales. *Entrelineas*, *3*(1), 52-64. https://doi.org/10.56368/Entrelineas315

Reyes, O., & Hernández-Moncada, M. (2019). *EMFIS, V-5.0* [*Estructura Metodológica Fundamentada para Investigaciones Sociales*]. https://doi.org/10.13140/RG.2.2.18861.13288

Rodríguez, C. G., Enríquez, M. R., & Cevallos, G. E. (2023). Gestión de operaciones y su impacto en la competitividad en la Comercializadora Dipor S.A. *Revista Tse'de*, *7*(2), 1-18. https://tsachila.edu.ec/ojs/index.php/TSEDE/issue/archive

Salinas, G. (2024). *Inteligencia artificial en la Gestión Administrativa*. LinkedIn.com. https://es.linkedin.com/pulse/inteligencia-artificial-en-la-gesti%C3%B3n-administrativa-german-salinas-bwkre

Sánchez, D. F., Philco, M. A., Salinas, J. M., & Pico, J. C. (2024). Impacto de la Inteligencia Artificial en la Precisión y Eficiencia de los Sistemas Contables Modernos. *Journal of Economic and Social Science Research*, *4*(3), 1–12. https://doi.org/10.55813/gaea/jessr/v4/n3/117

Sarmiento, M. G., Cerrito, A., Ruiz, A. M., Mora, C. P., G., Gómez, I. A., & Cervera, M. G. (2024). Intervención Organizacional bajo el enfoque de Ralph Kilmann: Percepción del clima organizacional. *Pistas Educativa*s, 148. https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/3742/2607

Soledispa, X. E., Pionce, J. M., & Sierra, M. C. (2022). La gestión administrativa, factor clave para la productividad y competitividad de las microempresas. *Ciencias Económicas y Empresariales*, *8*(1), 281-294. https://doi.org/10.23857/dc.v8i1.2571

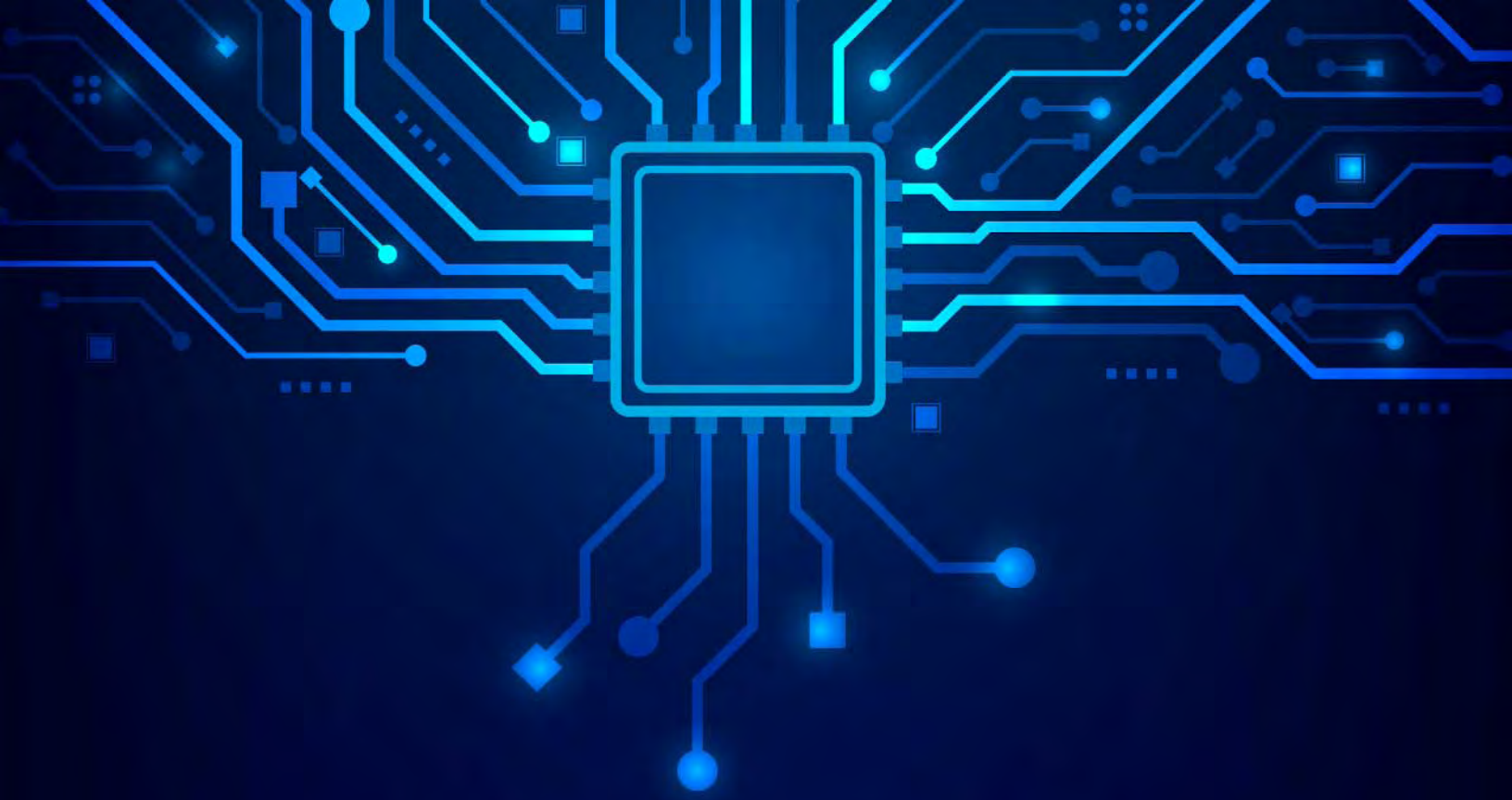
Soto, R. (2021). Eficiencia en la ejecución de proyectos de inversión. *Ciencia Latina Revista Científica Multidisciplinar*, *5*(2), 1726-1739. https://doi.org/10.37811/cl_rcm.v5i2.378

Teniente, E. V. (2026). Modelo de Gestión Administrativa apoyada en la Inteligencia Artificial, para microempresas comerciales del sector ferretero, establecidas en Celaya, Gto. México [Tesis Doctoral. Universidad del Centro del Bajío (UNICEBA)].

Terry, G. (1962). *Principios de Administración*. Editorial CECSA.

Tyler, C., Gordon, R., & Tyler, C. E., (2023). La administración empresarial y la utilización de la inteligencia artificial y GPT-4 aportes y desafíos para ingeniería de software y los sistemas de información. *Revista Científica Guacamaya*, *8*(1), 128-136. https://doi.org/10.48204/j.guacamaya.v8n1.a4323

Zabaleta, Y., Espinosa, R., & Benites, L. (2024). Deficiencias de control interno en las micro, pequeñas y medianas empresas: Caso municipio Celaya, Guanajuato, México. Scielo 18(1), http://scielo.sld.cu/scielo.php?pid=S2073-60612024000100008&script=sci_arttext&tlng=pt



21.

Modelos y aplicaciones de machine learning en la estratificación de riesgo clínico

Julia Dolores Toscano Garibay
Hospital Regional de Alta Especialidad de Ixtapaluca IMSS-Bienestar, México
ORCID: 0000-0001-5858-2814

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.21

21.

Modelos y aplicaciones del machine learning en la estratificación de riesgo clínico

Introducción

Los sistemas de salud a nivel mundial enfrentan retos derivados de la complejidad clínica en poblaciones cambiantes, los cuales demandan estrategias de priorización para optimizar la gestión de recursos (Rajkomar et al., 2019; Topol, 2019). Esta priorización, entendida como la capacidad de identificar proactivamente a los pacientes que más se beneficiarán de una atención personalizada, permite mejorar tiempos de respuesta, flujos de trabajo clínicos y asignaciones de personal, siendo aplicable en contextos tan distintos como el Servicio Nacional de Salud (NHS, por sus siglas en inglés) del Reino Unido, el sistema *Medicare* de Estados Unidos o redes de atención primaria en América Latina (Sendak et al., 2021; OECD, 2025). La inteligencia artificial (IA) y, en particular, el machine learning (ML) emergen actualmente como herramientas para analizar grandes volúmenes de datos clínicos y generar perfiles predictivos de pacientes que podrían guiar decisiones clínicas informadas y centradas en la persona (Davenport y Kalakota, 2019; Obermeyer et al., 2019).

La Organización Mundial de la Salud (WHO, por sus siglas en inglés) estima que 1,600 millones de personas en el mundo padecen multimorbilidad, incrementando la demanda de servicios incluso en contextos de alto ingreso, donde la Organización para la Cooperación y el Desarrollo Económicos (OCDE) documenta listas de espera superiores a 18 semanas en el NHS británico y demoras en atención especializada que superan los tres meses en Australia (OECD, 2025; WHO, 2025). Los *scores* tradicionales de estratificación clínica como *APACHE*, *SOFA* o *NEWS* tienen limitaciones al depender de un número reducido de variables y capturar solo interacciones lineales (Futoma et al., 2018; Sendak et al., 2021). El ML aborda estas limitaciones mediante algoritmos que aprenden patrones complejos directamente de los datos, con capacidad de procesamiento en tiempo real y de incorporar diversas fuentes de información (Choudhury et al., 2023; Wiens et al., 2019). Estudios con conjuntos de datos de hasta 42.9 millones de admisiones han demostrado que modelos

de ML alcanzan alta eficiencia (AUC > 0.89) en predicción de mortalidad intrahospitalaria (Rajkomar et al., 2018), y aplicaciones de aprendizaje por refuerzo han logrado mejoras del 17% en el pronóstico de demanda clínica (Schäfer et al., 2023).

El propósito de este capítulo es ofrecer una revisión conceptual de algunos modelos de ML aplicados en salud, con énfasis en su aplicabilidad universal en sistemas sanitarios de diversa escala y contexto. Se definen los fundamentos conceptuales de la IA y el ML, se describe la taxonomía, los principales tipos de modelos, se muestra su alcance documentado en investigación biomédica global y se exploran algunas aplicaciones en la estratificación de riesgo y priorización clínica.

Conceptos fundamentales

Inteligencia artificial como simulación de procesos cognitivos

La denominación *inteligencia artificial* fue acuñada formalmente en 1956 como la ciencia de construir máquinas capaces de emular comportamientos propios del razonamiento humano (Russell y Norvig, 2021). La IA abarca técnicas como el procesamiento del lenguaje natural, la visión artificial, los sistemas expertos y el ML (Davenport y Kalakota, 2019). En salud, su historia revela un movimiento desde sistemas expertos basados en reglas —como *MYCIN* en los años 70s— hasta modelos actuales de aprendizaje profundo capaces de detectar melanoma o interpretar electrocardiogramas con alta precisión (Esteva et al., 2017; Rajpurkar et al., 2022). La IA no reemplaza el juicio clínico, sino que lo amplifica, ofreciendo capacidad de procesamiento que aumenta el conocimiento individual (Topol, 2019).

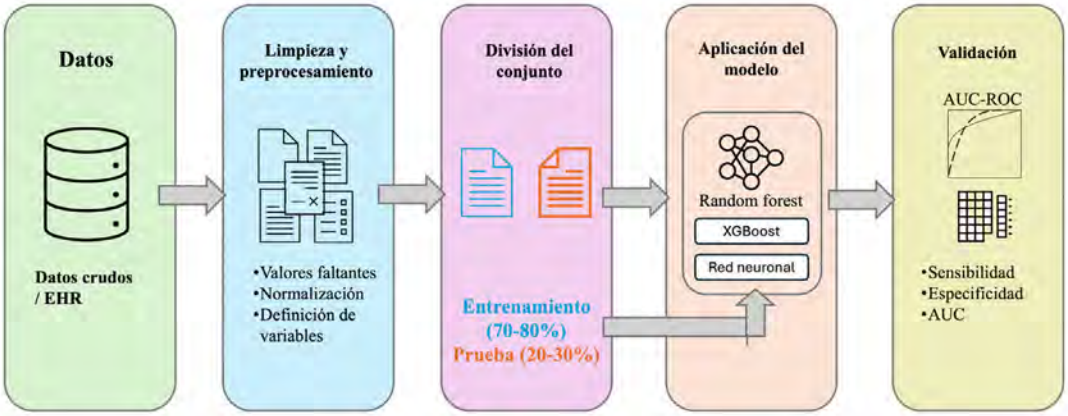
Machine learning: inteligencia artificial basada en aprendizaje de datos

El ML constituye una parte de la IA con uno de los más altos impactos en investigación biomédica, y se define como el conjunto de algoritmos computacionales que mejoran el desempeño en tareas específicas mediante la exposición reiterativa a datos, con lo cual no se requiere la programación explícita de reglas para cada caso particular (Beam y Kohane, 2018; Deo, 2015). A diferencia de los sistemas expertos clásicos, un modelo de ML extrae automáticamente las relaciones estadísticas relevantes de un conjunto de datos de entrenamiento, generando representaciones internas que le permiten generalizar a casos no vistos previamente (conjunto de prueba) (Obermeyer y Emanuel, 2016; Choudhury et al., 2023). Un flujo de trabajo típico en las técnicas de ML se muestra en la Figura 1. Los datos crudos son

organizados, normalizados y divididos para la generación del modelo matemático. Cuando este último es aplicado al subconjunto de prueba, los resultados muestran patrones predictivos aplicables a otras poblaciones. Esta capacidad de aprendizaje inductivo es lo que confiere al ML su flexibilidad para abordar problemas clínicos de alta complejidad, donde las interacciones entre variables son no lineales y multidimensionales (Saria et al., 2019).

Figura 1

Flujo de trabajo del machine learning: del dato crudo a la validación del modelo



Nota. Elaboración propia con asistencia de inteligencia artificial (Perplexity AI, 2026).

Es fundamental distinguir el ML de los métodos estadísticos tradicionales. Mientras que la estadística inferencial busca cuantificar relaciones entre variables para formular conclusiones causales, el ML se orienta principalmente hacia la maximización de la precisión predictiva, aceptando modelos de alta complejidad no lineal (Steyerberg et al., 2019; Saria et al., 2019).

La taxonomía del ML se organiza en tres tipos de aprendizaje. El aprendizaje supervisado entrena algoritmos sobre datos etiquetados —con una variable de resultado conocida— para tareas de clasificación y regresión. El aprendizaje no supervisado trabaja sobre datos sin etiquetas para descubrir estructuras latentes como agrupamientos, útil para la caracterización clínica de los pacientes (fenotipado). El aprendizaje por refuerzo opera mediante un agente que interactúa con un entorno y aprende a seleccionar acciones que maximizan una recompensa acumulada, aplicable a problemas de optimización secuencial (Gottesman et al., 2019; Petch et al., 2022).

Taxonomía de Modelos de Machine Learning

Modelos supervisados

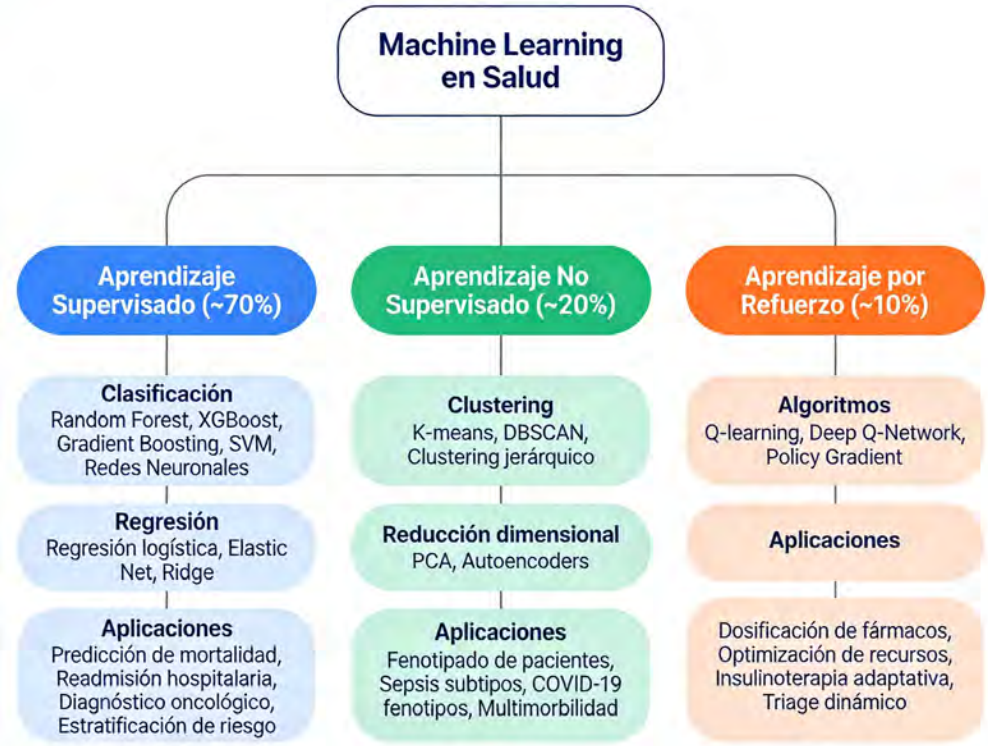
Los modelos de aprendizaje supervisado aprenden una función que rastrea la relación entre variables de entrada (características del paciente) y una variable de salida previamente etiquetada, como por ejemplo, la probabilidad de readmisión o el riesgo de mortalidad (Deo, 2015; Beam y Kohane, 2018). La Figura 2 muestra una síntesis de la taxonomía general de los modelos ML que serán descritos a continuación. La regresión logística ha sido el estándar de referencia por su interpretabilidad mediante cocientes de momios, pero el supuesto de linealidad inherente a este algoritmo limita su capacidad para capturar interacciones complejas (Saria et al., 2019). Los árboles de decisión superan esta limitación al dividir características de manera recursiva y no lineal, generando reglas visualmente interpretables, aunque son propensos al denominado sobreajuste (Deo, 2015). Las máquinas de vectores de soporte (SVM) trazan la frontera matemática más eficiente para separar categorías clínicas, con capacidad de operar con datos complejos y multivariables (Petch et al., 2022). Los métodos de ensamble constituyen el estado del arte en predicción sobre datos tabulares: el random forest agrega predicciones de múltiples árboles reduciendo la varianza, mientras que gradient boosting y XGBoost entrenan secuencialmente árboles que corrigen errores del árbol anterior, logrando AUC sostenidos superiores a 0.90 en predicción de mortalidad y falla renal aguda (Molina et al., 2023; Rashid et al., 2023).

Modelos no supervisados

Los algoritmos de aprendizaje no supervisado operan sobre datos sin etiquetas y descubren estructuras internas como agrupamientos, útiles para el fenotipado de pacientes: identificar subgrupos con patrones similares de comorbilidades, utilización de servicios o respuesta a tratamientos (Obermeyer et al., 2019; Loftus et al., 2022). El algoritmo K-means es el método de clustering más utilizado, que divide el conjunto en k grupos cuya formación minimiza la varianza intragrupo. Su aplicación en pacientes con COVID-19 identificó cinco fenotipos clínicamente distintos con diferencias significativas en mortalidad y respuesta a ventilación mecánica (Palomar-Lever et al., 2021). El análisis de componentes principales (PCA) y otros métodos como t-SNE y UMAP condensan cientos de variables clínicas en sus componentes esenciales, conservando las semejanzas entre pacientes y haciendo manejable su agrupación, facilitando la visualización y el mejoramiento del rendimiento de modelos predictivos subsecuentes (Beam & Kohane, 2018).

Figura 2

Taxonomía de algoritmos de machine learning y sus aplicaciones en salud



Nota. Elaboración propia con asistencia de inteligencia artificial (Perplexity AI, 2026).

Aprendizaje por refuerzo

El aprendizaje por refuerzo (AR) difiere de los enfoques supervisados y no supervisados: un agente aprende mediante la interacción iterativa con un entorno dinámico, tomando acciones que generan recompensas y ajustando su política de decisión para maximizar la recompensa acumulada a lo largo del tiempo (Gottesman et al., 2019; Petch et al., 2022). Esta arquitectura resulta adecuada para problemas clínicos de decisión secuencial, como la titulación de vasopresores en sepsis o la dosificación de heparina en anticoagulación. Políticas de manejo de líquidos y vasopresores generadas mediante AR para pacientes con sepsis redujeron la mortalidad relativa en 10.23% frente a la práctica clínica estándar en estudios retrospectivos a gran escala (Gottesman et al., 2019). Por otro lado, el aprendizaje semisupervisado combina datos etiquetados y no etiquetados para mejorar el rendimiento cuando el etiquetado es costoso, especialmente relevante en medicina donde la anotación por expertos es limitada.

Deep learning e híbridos

El *deep learning* (DL) se distingue por el uso de redes neuronales artificiales con múltiples capas capaces de extraer automáticamente características de alta abstracción a partir de datos no estructurados como imágenes, señales temporales y texto clínico (Rajpurkar et al., 2022; Topol, 2019). Las arquitecturas más relevantes de DL incluyen las redes neuronales convolucionales (CNN) en análisis de imágenes médicas con exactitudes de 95-98%, las redes neuronales recurrentes (RNN/LSTM) para series temporales de signos vitales, y los transformers en procesamiento del lenguaje natural aplicado a notas clínicas (Jiang et al., 2024). Los modelos de lenguaje de gran escala (LLM) como *GPT-4* y *Med-PaLM 2* han demostrado capacidad para responder preguntas médicas complejas y asistir en diagnóstico diferencial, con implicaciones para automatizar el procesamiento de información no estructurada (Jiang et al., 2024). Los modelos híbridos combinan dos o más paradigmas de ML para capitalizar las fortalezas de cada enfoque, como ensambles de redes neuronales con SVM o XGBoost con regresión de Cox, superando en capacidad predictiva a cada componente por separado (Molina et al., 2023).

Ejemplos de aplicaciones en salud del machine learning

La transición desde la investigación hacia el entorno clínico ha consolidado al ML como una herramienta útil en prácticamente todas las especialidades médicas. La magnitud y diversidad de esta producción científica —que comprende miles de publicaciones, revisiones sistemáticas y meta-análisis durante la última década— exige una lectura organizada que permita comprender cuántos y cuáles tipos de modelos han sido efectivamente aplicados en salud, en qué proporciones se distribuyen entre los tipos de aprendizaje, y qué dominios clínicos han concentrado el mayor volumen de investigación y evidencia de resultados (Choudhury et al., 2023; Rajpurkar et al., 2022).

La Tabla 1 sintetiza la distribución de aplicaciones de ML en salud por tipo de aprendizaje. Los modelos supervisados representan aproximadamente el 70% de los estudios, concentrándose en tareas de predicción clínica donde una variable de resultado conocida —mortalidad, readmisión, diagnóstico— sirve como señal de supervisión para el entrenamiento algorítmico. Los modelos no supervisados, que representan el 20%, se aplican al fenotipado computacional: la identificación de subgrupos de pacientes con perfiles similares sin criterios diagnósticos predefinidos, revelando heterogeneidades clínicas con implicaciones para la personalización del tratamiento. El aprendizaje por refuerzo, con el 10% de las

aplicaciones, se especializa en optimización de decisiones secuenciales en entornos dinámicos, como la dosificación adaptativa de fármacos o la gestión de recursos hospitalarios.

Tabla 1
Distribución de aplicaciones de machine learning en salud según tipo de aprendizaje

Aprendizaje	Proporción	Algoritmos representativos	Aplicaciones clínicas documentadas	Evidencia clave
Supervisado	~70%	Random Forest, XGBoost, Gradient Boosting, Redes neuronales, SVM	Predicción de mortalidad (AUC 0.86), readmisión hospitalaria, diagnóstico oncológico, estratificación de riesgo cardiovascular	Delpino et al. (2025); Rashid et al. (2023); Ghasemi et al. (2025)
No supervisado	~20%	K-means, DBSCAN, Clustering jerárquico, PCA, Autoencoders	Fenotipado de sepsis (2-5 subtipos), fenotipos COVID-19 (mortalidad 0%-10.6%), farmacofenotipos en UCI, trasplante renal	Nalinthasnai et al. (2025); Li et al. (2025); Sikora et al. (2023)
Aprendizaje por refuerzo	~10%	Q-learning, Deep Q-Network, Policy Gradient	Dosificación de vasopresores/ antibióticos, insulinoterapia adaptativa (TIR superior), radioterapia adaptativa, gestión de camas	Coronato et al. (2020); Nambiar et al. (2024); Schäfer et al. (2023)

Nota. Elaboración propia con base en revisiones sistemáticas citadas. TIR = tiempo en rango glucémico; AUC = área bajo la curva ROC; UCI = unidad de cuidados intensivos.

Machine learning para perfiles específicos y priorización

La literatura acumulada durante la última década evidencia que los modelos de ML no solo predicen eventos clínicos con mayor precisión que los métodos convencionales, sino que estructuran la información disponible en los registros electrónicos de salud de manera que permite identificar, antes de que se manifieste el deterioro clínico, qué pacientes requieren

mayor intensidad de atención, seguimiento proactivo, o derivación prioritaria a servicios especializados (Sendak et al., 2021; Rajpurkar et al., 2022). Esta capacidad de anticipación fundamenta su valor en cuatro grandes aplicaciones de priorización que se sintetizaron en la Tabla 2: estratificación de riesgo, apoyo al triage en urgencias, predicción de alta demanda y la integración operativa en sistemas clínicos de apoyo a la decisión en tiempo real.

La estratificación de riesgo mediante ML asigna a cada paciente una puntuación a partir de múltiples variables clínicas, administrativas y sociodemográficas, capturando interacciones no lineales que los scores tradicionales no pueden modelar. El triage en urgencias se transforma de un proceso binario basado en síntomas a uno continuo de estimación de riesgo individualizado, con modelos como eCART superando consistentemente a los scores convencionales. La predicción de alta demanda anticipa picos en el volumen de pacientes, ocupación de camas y consumo de recursos, permitiendo medidas preventivas como redistribución de personal o habilitación de capacidad adicional. Finalmente, la integración operativa en sistemas clínicos constituye el mayor desafío traslacional: solo el 5% de los modelos de ML desarrollados alcanzan implementación clínica real, evidenciando una brecha profunda entre investigación y adopción práctica (Choudhury et al., 2023; Preti et al., 2024).

Tabla 2

Aplicaciones de machine learning para priorización clínica y gestión de recursos en salud

Aplicación	Función principal	Modelos de mayor rendimiento	Resultados documentados	Desafíos
Estratificación de riesgo	Asignar puntuación de riesgo individual para orientar seguimiento e intervención preventiva	Gradient Boosting (RMSE 1.48, R^2=0.81), Random Forest, Redes neuronales	Exactitud 78.39% vs. 47.70% del score LACE+; reducción del 48% en readmisiones con modelo Re-Admit	Brecha traslacional: solo 5% de modelos alcanza implementación clínica
Triage y urgencias	Estimar riesgo de deterioro en tiempo real para priorizar atención en urgencias	eCART (Gradient Boosting), BioClinicalBERT (NLP), XGBoost	SERP superó a NEWS, MEWS y MEWS2; precisión 0.92 con notas de triage vs. 0.12 para NEWS	Variabilidad en rendimiento entre poblaciones; riesgo de sobreconfianza
Predicción de alta demanda	Anticipar volumen de pacientes, ocupación de camas y consumo de recursos	XGBoost, LSTM, SVM-kernel (K-SVR)	MAPE 8.1% en ocupación hospitalaria; superioridad sobre ARIMA en patrones estacionales	Dependencia de datos históricos de calidad; variabilidad contextual

Tabla 2
Aplicaciones de machine learning para priorización clínica y gestión de recursos en salud

Aplicación	Función principal	Modelos de mayor rendimiento	Resultados documentados	Desafíos
Integración en sistemas clínicos	Operacionalizar modelos predictivos dentro del flujo de trabajo clínico del EHR	CDSS basados en HL7 FHIR; módulos no interruptivos	Mayor impacto clínico cuando integrado en EHR vs. herramienta externa	Interoperabilidad limitada; ausencia de marcos regulatorios nacionales

Nota. EHR = historia clínica electrónica; NLP = procesamiento de lenguaje natural; MAPE = error absoluto porcentual medio; CDSS = sistema de apoyo a la decisión clínica. Choudhury et al. (2023); Preti et al. (2024); Watson et al. (2024).

Discusión y conclusiones

El recorrido de conceptos realizado a lo largo de este capítulo permite afirmar que el machine learning representa un cambio cualitativo sustancial en la capacidad de los sistemas de salud para comprender, anticipar y responder a las necesidades de sus poblaciones. La revisión de la taxonomía y las aplicaciones documentadas evidencia que esta transformación no es un horizonte tecnológico lejano, sino una realidad operativa en progreso, con múltiples modelos validados, herramientas integradas en flujos de trabajo clínico real y resultados reproducibles en múltiples contextos geográficos e institucionales (Choudhury et al., 2023; Rajpurkar et al., 2022).

Retos pendientes: interpretabilidad, equidad y marco normativo

La interpretabilidad de los algoritmos constituye un reto central cuando se busca adoptar ML en entornos donde la justificación de decisiones clínicas es esencial para su aceptación por el profesional de la salud y el afianzamiento de la confianza del paciente (Saleem et al., 2025). Por ejemplo, aunque los modelos de ensamble y las redes neuronales profundas superan en precisión a la regresión logística, su opacidad técnica dificulta entender por qué un paciente es clasificado como de alto riesgo. Técnicas como SHAP (SHapley Additive exPlanations) y LIME (Local Interpretable Model-agnostic Explanations) emergieron entonces para generar explicaciones post-hoc de tales modelos complejos. Saleem et al. (2025) mostraron en su estudio aplicando SHAP a un modelo de gradient boosting para predicción de readmisión hospitalaria que las variables más influyentes fueron

el número de diagnósticos previos, la duración de la estancia y el nivel de albúmina sérica, información que resultó accionable para el equipo clínico al diseñar intervenciones preventivas específicas. Por lo cual, esta aproximación debe ser generalizada en las aplicaciones de las técnicas de ML.

El sesgo algorítmico representa otro obstáculo ético de primera magnitud, documentado en el estudio seminal de Obermeyer et al. (2019), que demostró que un modelo comercial de predicción de riesgo ampliamente utilizado en hospitales estadounidenses subestimaba sistemáticamente la necesidad de atención de pacientes afroamericanos, quienes a igual puntuación de riesgo que pacientes blancos presentaban mayor carga de enfermedad crónica no reconocida. Este sesgo derivaba del uso del costo de atención como *proxy* de necesidad de atención, una decisión de diseño que perpetúa desigualdades estructurales preexistentes (Obermeyer et al., 2019). La equidad algorítmica en consecuencia requiere auditorías sistemáticas del rendimiento desagregado por subgrupos vulnerables, el uso de métricas de equidad durante el desarrollo y validación, y la participación de comunidades afectadas en el diseño de sistemas (Norgeot et al., 2019).

El marco regulatorio de modelos de ML en salud demanda una gestión normativa clara que equilibre la innovación con la protección de los pacientes. La actualización continua de modelos para mantener su rendimiento ante cambios epidemiológicos (*concept drift*) requiere procesos de validación prospectiva continua, que son costosos y logísticamente complejos (Wiens et al., 2019). La interoperabilidad entre sistemas de información hospitalaria heterogéneos y la ausencia de estándares unificados de documentación de modelos de ML dificultan la transferibilidad de herramientas entre instituciones (Lee et al., 2022). Finalmente, la educación de profesionales de la salud en competencias de alfabetización digital y pensamiento crítico sobre las salidas de modelos de ML es indispensable para una integración responsable que preserve la autonomía clínica y la centralidad del paciente (Topol, 2019).

Sin embargo, el machine learning representa una de las herramientas más promisorias para transformar la investigación en salud, con aplicaciones documentadas que mejoran tanto la precisión clínica como la eficiencia operativa de los sistemas sanitarios. Lo que resta no es demostrar su utilidad, sino construir los andamiajes éticos, regulatorios e institucionales que permitan que esa utilidad llegue con equidad a todos los pacientes. Su consolidación como práctica estándar exige resolver los desafíos estructurales mencionados anteriormente que limitan aún su transferibilidad a entornos clínicos reales. Avanzar en ambas dimensiones —técnica e institucional— es la condición necesaria para que el ML cumpla su promesa de

una medicina más precisa, equitativa y centrada en la persona. El camino está trazado; el reto es recorrerlo con responsabilidad.

Referencias

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Annual Review of Medicine*, 69, 153-161. https://doi.org/10.1146/annurev-med-050516-033destress

Choudhury, A., Asan, O., & Hu, Q. (2023). Systematic reviews of machine learning in healthcare: A literature review. *Journal of Medical Internet Research*, 25, e46089. https://doi.org/10.2196/46089

Coronato, A., Naeem, M., De Pietro, G., & Paragliola, G. (2020). Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine*, 109, 101964. https://doi.org/10.1016/j.artmed.2020.101964

Davenport, T. y Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94-98. https://doi.org/10.7861/futurehosp.6-2-94

Delpino, F. M., Flores, T. R., Portela, M. A. P., Hartwig, F. P., Menezes, A. M. B., Barcelos, R. K., & Thumé, E. (2025). Global performance of machine learning models to predict all-cause mortality. *Scientific Reports*, *15*(1), 10381. https://doi.org/10.1038/s41598-025-94576-5

Deo, R. C. (2015). Machine learning in medicine. *Circulation*, *132*(20), 1920-1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://doi.org/10.1038/nature21056

Futoma, J., Hariharan, S., & Heller, K. (2018). Learning to detect sepsis with a multitask Gaussian process RNN classifier. *JAMIA Open*, *1*(1), 77-85. https://doi.org/10.1093/jamiaopen/ooy014

Ghasemi, N., Haghighi, M. S., Dehghani, S., & Baradaran, H. R. (2025). Artificial intelligence in breast cancer survival prediction: A comprehensive systematic review and meta-analysis. *Frontiers in Oncology*, 14, 1420328. https://doi.org/10.3389/fonc.2024.1420328

Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., & Celi, L. A. (2019). Guidelines for reinforcement learning in healthcare. *Nature Medicine*, *25*(1), 16-18. https://doi.org/10.1038/s41591-018-0310-5

Jiang, L. Y., Liu, X. C., Nejatian, N. P., Nasir-Moin, M., Wang, D., Abidin, A., Eaton, K., Riina, H. A., Laufer, I., Punjabi, P., Miceli, M., Kim, N. J., Orillac, C., Zanos, P., Kenigsberg, E., Costa, A. B., Mirhaji, P., Fattah, M., Moll, M. K. ... Narayan, J. (2024). Health system-scale language models are all-purpose prediction engines. *Nature*, *631*(8020), 357-365. https://doi.org/10.1038/s41586-024-07603-2

Lee, S., Lee, S. H., & Kim, N. (2022). Development of an interoperable and easily transferable clinical decision support system deployment platform: System design and development study. *Journal of Medical Internet Research*, *24*(7), e37214. https://doi.org/10.2196/37214

Li, N., Sheng, B., Mao, Y., Li, T., Chen, J., Xu, Y., & Li, C. (2025). Unsupervised clustering for sepsis identification in large-scale epidemiologic surveillance data. *Intensive Care Medicine Experimental*, *13*(1), 24. https://doi.org/10.1186/s40635-025-00726-2

Loftus, T. J., Shickel, B., Balch, J. A., Upchurch, G. R. y Bihorac, A. (2022). Phenotype clustering in health care: A narrative review of methods and applications. *npj Digital Medicine*, *5*(1), 105. https://doi.org/10.1038/s41746-022-00648-9

Molina, A., Puig-Domingo, M., Lecube, A., Simó-Servat, O., & Ciudin, A. (2023). Development and validation of an XGBoost-algorithm-powered survival model for predicting in-hospital mortality based on 545,388 isolated severe traumatic brain injury patients from the TQIP database. *Frontiers in Neurology*, 14, 1246886. https://doi.org/10.3389/fneur.2023.1246886

Nalinthasnai, N., Sirilert, S., & Tongsong, T. (2025). Unsupervised machine learning clustering approach for phenotyping and clinical outcomes in patients with severe COVID-19. *Frontiers in Medicine*, 12, 1543279. https://doi.org/10.3389/fmed.2025.1543279

Nambiar, M., Wilson, B. M., Bhatt, D. L., & Pfeifer, J. M. (2024). Reinforcement learning in healthcare: Applications for personalized treatment planning and clinical decision support. *Journal of Artificial Intelligence and Machine Learning*, *1*(1), 10. https://doi.org/10.46965/jaim.1.1.10

Norgeot, B., Glicksberg, B. S., & Butte, A. J. (2019). A call for deep-learning healthcare. *Nature Medicine*, *25*(1), 14-15. https://doi.org/10.1038/s41591-018-0320-3

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future: Big data, machine learning, and clinical medicine. *New England Journal of Medicine, 375*(13), 1216-1219. https://doi.org/10.1056/NEJMp1606181

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

OECD. (2025). T*he future of health systems*. OECD. https://www.oecd.org/en/topics/policy-issues/the-future-of-health-systems.html

Palomar-Lever, A., García-Gutiérrez, S., Prieto, D., González-García, B., Cruz-Gómez, A. J., & Merino, J. (2021). Deploying unsupervised clustering analysis to derive clinical phenotypes and risk factors associated with mortality risk in 2022 critically ill COVID-19 patients in Spain. *Critical Care*, *25*(1), 63. https://doi.org/10.1186/s13054-021-03487-8

Petch, J., Di, S., & Nelson, W. (2022). Opening the black box: The promise and limitations of explainable machine learning in cardiology. *Canadian Journal of Cardiology*, *38*(2), 204-213. https://doi.org/10.1016/j.cjca.2021.09.004

Preti, L. M., Elkin, P. L., Topaz, M., & Sridharan, S. (2024). Implementation of machine learning applications in health care settings: Systematic review. *Journal of Medical Internet Research*, 26, e55897. https://doi.org/10.2196/55897

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., ... Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1, 18. https://doi.org/10.1038/s41746-018-0029-1

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31-38. https://doi.org/10.1038/s41591-021-01614-0

Rashid, M. M., Liu, J., & Yi, C. (2023). Machine learning based readmission and mortality prediction in heart failure patients. *Scientific Reports*, *13*(1), 18476. https://doi.org/10.1038/s41598-023-45893-0

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approac*h (4.ª ed.). Pearson.

Saleem, J. A., Siek, K. A., Shabbir, S. A., & Garg, V. (2025). Comparison of SHAP and clinician friendly explanations reveals effects on clinical decision behaviour. *npj Digital Medicine*, *8*(1), 467. https://doi.org/10.1038/s41746-025-01467-y

Saria, S., Butte, A., & Sheikh, A. (2019). Better medicine through machine learning: What's real, and what's artificial? *PLOS Medicine*, *15*(12), e1002721. https://doi.org/10.1371/journal.pmed.1002721

Schäfer, F., Schiffer, M., & Walther, G. (2023). Data-driven hospital staff and resources allocation using agent-based simulation and deep reinforcement learning. *Engineering Applications of Artificial Intelligence*, 126, 107097. https://doi.org/10.1016/j.engappai.2023.107097

Sendak, M. P., Elish, M. C., Gao, M., Futoma, J., Ratliff, W., Nichols, M., Bedoya, A., Balu, S., & O'Brien, C. (2021). Machine learning for patient risk stratification: Standing on, or looking over, the shoulders of clinicians? *npj Digital Medicine*, *4*(1), 56. https://doi.org/10.1038/s41746-021-00426-3

Sikora, A., Zimmerman, L., Mueller, S. W., Bhatt, D. L., & Rech, M. A. (2023). Cluster analysis driven by unsupervised latent feature learning of ICU medications. *Scientific Reports*, *13*(1), 15612. https://doi.org/10.1038/s41598-023-42617-y

Steyerberg, E. W., Moons, K. G. M., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., Riley, R. D., Hemingway, H., & Altman, D. G. (2019). Prognosis research strategy (PROGRESS) 3: Prognostic model research. *PLOS Medicine*, *10*(2), e1001381. https://doi.org/10.1371/journal.pmed.1001381

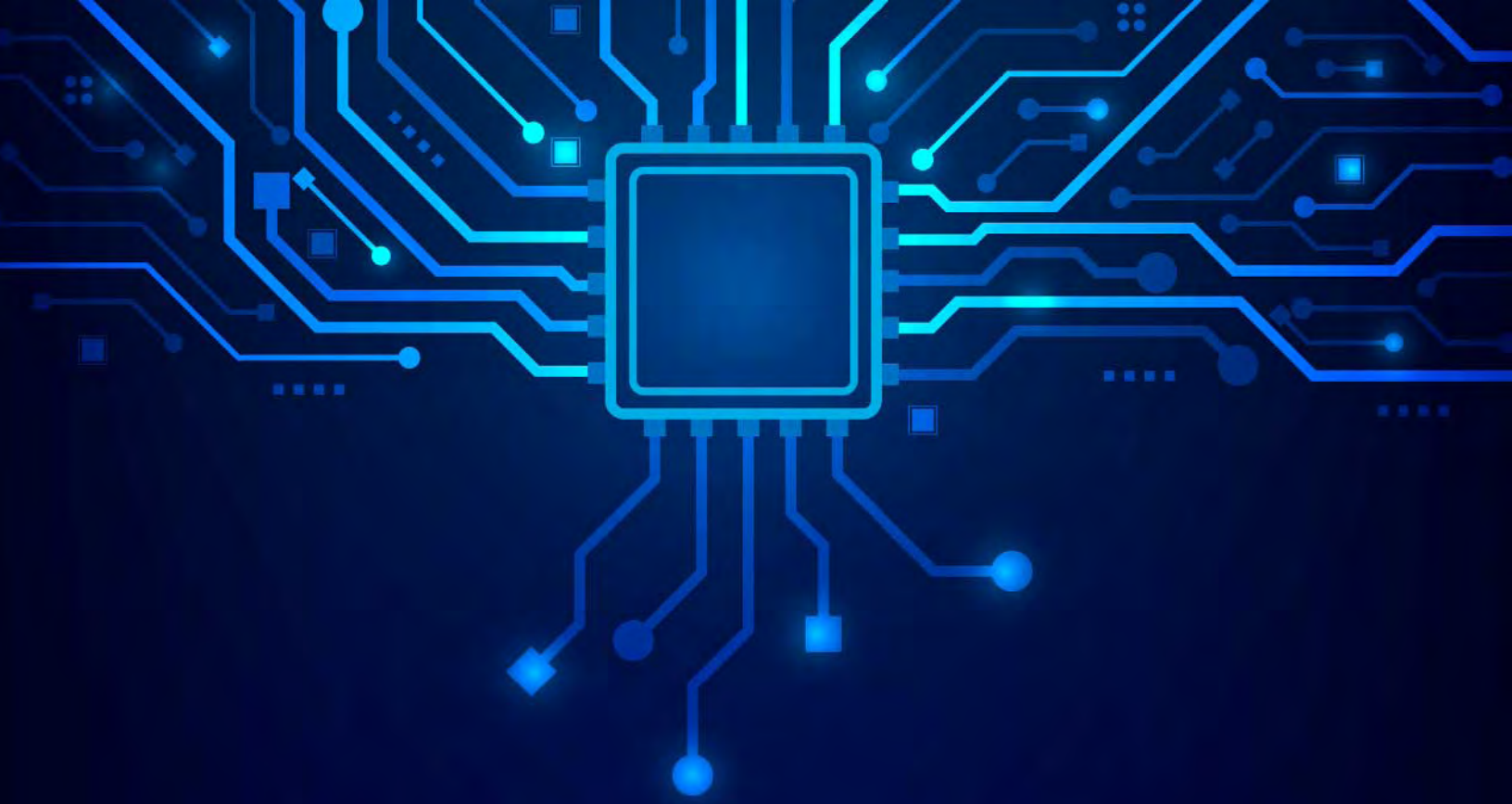
Tariku, M. F., Kuru, K., & Okolo, C. (2025). Predicting total healthcare demand using machine learning: Separate and combined analysis of predisposing, enabling, and need factors. *BMC Health Services Research*, *25*(1), 401. https://doi.org/10.1186/s12913-025-12588-x

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Watson, M., Ireland, G., Stirling, A., Lawrenson, P., Chuter, A., Dobie, G., Dent, J., & Mair, F. S. (2024). Performance of machine learning versus the National Early Warning Score for predicting patient deterioration risk: A single-site study of emergency admissions. *BMJ Open*, *14*(12), e085211. https://doi.org/10.1136/bmjopen-2024-085211

WHO. (2025). *Countries are already experiencing significant health system disruptions*. WHO. https://www.who.int/news/item/10-04-2025-countries-are-already-experiencing-significant-health-system-disruptions—-who

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *npj Digital Medicine*, 2, 138. https://doi.org/10.1038/s41746-019-0154-8



22.

Entre principios y práctica: revisión de marcos regulatorios y éticos sobre inteligencia artificial en la investigación científica

Miguel Ángel Medina Romero
Universidad Michoacana de San Nicolás de Hidalgo, México
ORCID: 0000-0003-4067-2816

Tania Haidée Torres Chávez
Universidad Michoacana de San Nicolás de Hidalgo, México
ORCID: 0009-0000-6040-3870

Rodrigo Ochoa Figueroa
Universidad Michoacana de San Nicolás de Hidalgo, México
ORCID: 0000-0003-1957-8147

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.22

22.

Entre principios y práctica: revisión de marcos regulatorios y éticos sobre inteligencia artificial en la investigación científica

Introducción

La rápida expansión de la inteligencia artificial (IA) en la investigación científica ha reconfigurado de manera profunda los procesos de producción de conocimiento, desde la obtención y análisis de datos hasta la redacción y difusión de resultados, generando beneficios en términos de eficiencia, escala y capacidad analítica, pero también nuevos riesgos éticos y epistémicos que impactan la objetividad, la transparencia, la rendición de cuentas y la confianza pública en la ciencia (Resnik & Hosseini, 2025; Chen et al., 2024; Limongi, 2024). En este contexto, la ética de la ciencia y la integridad de la investigación siguen constituyendo el núcleo normativo para evaluar el uso responsable de tecnologías emergentes: principios clásicos como honestidad, transparencia y responsabilidad conservan su vigencia, pero requieren ser reinterpretados ante desafíos específicos vinculados con la opacidad algorítmica, los sesgos, la trazabilidad de los procesos de inferencia, la autoría y la aparición de nuevas formas de mala conducta académica asociadas al uso de sistemas de IA (Resnik & Hosseini, 2025; Chen et al., 2024).

Paralelamente, han cobrado fuerza marcos de ética de la IA y de gobernanza responsable impulsados por organismos internacionales. La *Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO* —Recomendación de la UNESCO, en adelante— establece principios como derechos humanos, justicia, inclusividad, transparencia y rendición de cuentas, e insiste en la necesidad de marcos normativos multinivel que articulen estándares internacionales con adaptaciones nacionales e institucionales, incluyendo explícitamente el ámbito de la investigación (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2021). En la misma dirección, las directrices de la Comisión Europea para el uso responsable de IA generativa en la investigación proponen principios

operativos que vinculan la integridad científica con la protección de datos y privacidad, la mitigación de sesgos, la transparencia en la comunicación del uso de IA y la evaluación del impacto social, ofreciendo un marco de referencia para políticas internas de las organizaciones de investigación (European Commission & European Research Area Forum, 2024). Sobre esta base, contribuciones recientes sistematizan "principios nucleares" para el uso responsable de IA generativa en investigación —regulación, seguridad de datos, control de calidad, originalidad, responsabilidad, transparencia e impacto— y subrayan la necesidad de implementarlos de forma coherente en los planos internacional, nacional e institucional, configurando un entramado de normas jurídicas, lineamientos éticos y políticas organizacionales que debería garantizar prácticas científicas íntegras y socialmente responsables (Knöchel et al., 2025; Limongi, 2024).

No obstante, la literatura empírica y conceptual disponible muestra varios vacíos: aunque identifica riesgos, formula principios y ofrece recomendaciones generales, tiende a tratar la ciencia como un todo homogéneo, sin diferenciar sistemáticamente entre disciplinas ni mapear de manera comparativa los marcos regulatorios y lineamientos éticos existentes según su nivel y su atención específica a las ciencias sociales y las humanidades. Esto dificulta conocer con precisión qué normas están vigentes, cómo se distribuyen entre niveles internacional, nacional e institucional, qué grado de coherencia guardan entre sí y hasta qué punto responden a las particularidades éticas y metodológicas de las ciencias sociales y las humanidades, caracterizadas por el trabajo con datos sensibles, relaciones cercanas con participantes y métodos interpretativos. Frente a este panorama, se justifica la necesidad de un estudio que identifique y caracterice sistemáticamente estos marcos, y que evalúe de forma comparativa su capacidad para orientar un uso responsable de la IA en la investigación científica, con especial énfasis en las ciencias sociales y las humanidades.

Método

Este estudio se diseñó como una revisión sistemática de literatura científica y de documentos normativos y de política pública, siguiendo las directrices PRISMA 2020 para asegurar transparencia, exhaustividad y reproducibilidad en todas las etapas (Page et al., 2021). El objetivo fue identificar y caracterizar marcos regulatorios y lineamientos éticos sobre el uso de inteligencia artificial en la investigación científica entre 2019 y 2025, describiendo su alcance y nivel (internacional, nacional e institucional) y valorando en qué medida incorporan disposiciones explícitas y diferenciadas para las ciencias sociales y las humanidades (Her-

nández-Sampieri & Mendoza, 2018). El proceso se desarrolló en cuatro fases: identificación, cribado, elegibilidad e inclusión. En la identificación se realizaron búsquedas sistemáticas en bases de datos como *Scopus*, *Web of Science* y *Google Scholar*, y en portales de organismos internacionales y agencias públicas, especialmente UNESCO y la Comisión Europea, así como de agencias nacionales de ciencia, comités de ética y universidades (UNESCO, 2021; European Commission & European Research Area Forum, 2024). Los registros se gestionaron en un gestor bibliográfico, se eliminaron duplicados y, en el cribado, dos revisores evaluaron de forma independiente títulos y resúmenes; los textos potencialmente relevantes se analizaron a texto completo para determinar su elegibilidad y conformar el corpus final, documentado en un diagrama de flujo PRISMA (Moher, 2029).

Las preguntas de investigación se definieron para operacionalizar el objetivo general: identificar qué marcos y lineamientos sobre IA en la investigación se emitieron entre 2019 y 2025 a nivel internacional, nacional e institucional; describir su alcance y contenido en las distintas fases del proceso de investigación; y examinar hasta qué punto incorporan disposiciones específicas para las ciencias sociales y las humanidades, en diálogo con los debates sobre ética, integridad y uso responsable de IA generativa (Limongi, 2024; Chen et al., 2024; Knöchel et al., 2025; Medina-Romero et al., 2023). La estrategia de búsqueda combinó términos en español e inglés relativos a inteligencia artificial, investigación científica, regulación, ética y ciencias sociales y humanidades, incluyendo expresiones como “inteligencia artificial”, “artificial intelligence”, “generative AI”, “research integrity”, “regulation”, “governance”, “ethical guidelines”, “social sciences”, “humanities” y “SSH”. Además, se realizaron búsquedas dirigidas de documentos normativos clave, como la Recomendación de la UNESCO sobre ética de la IA y las directrices europeas para el uso responsable de IA generativa en la investigación, dada su centralidad en la definición de principios de integridad y gobernanza de la IA en ciencia (UNESCO, 2021; European Commission & European Research Area Forum, 2024; Knöchel et al., 2025; Limongi, 2024).

Los criterios de inclusión, coherentes con el objetivo del estudio y con estándares recientes sobre integridad y gobernanza de la IA, contemplaron documentos publicados o adoptados entre el 1 de enero de 2019 y el 31 de diciembre de 2025, en inglés o español, que abordaran explícitamente el uso de IA en la investigación científica y/o marcos de regulación, gobernanza o ética aplicados a IA en contextos de investigación (Chen et al., 2024; Limongi, 2024; Medina-Romero, 2025). Se incluyeron artículos empíricos, teóricos y de análisis normativo revisados por pares, así como recomendaciones, códigos de conducta,

directrices y políticas institucionales emitidos por organismos internacionales, agencias nacionales, comités de ética y universidades. Se exigió que los documentos permitieran identificar el nivel del marco, su alcance (general o disciplinar) y la presencia o ausencia de disposiciones explícitas para ciencias sociales y humanidades. Se excluyeron textos fuera del periodo, en otros idiomas, trabajos centrados solo en desarrollos técnicos o aplicaciones comerciales sin vínculo con la investigación académica, documentos normativos sin referencia a la investigación, materiales no revisados por pares y documentos institucionales sin carácter directivo claro (Page et al., 2021).

A partir del corpus final se realizó una extracción sistemática de datos mediante una matriz diseñada para registrar tipo de documento, año, origen, nivel, ámbito, referencia explícita a la investigación y presencia de disposiciones específicas para las ciencias sociales y las humanidades. El análisis se llevó a cabo mediante síntesis cualitativa comparativa: primero se elaboró un mapa descriptivo de marcos y lineamientos por nivel y tipo; después se aplicó un análisis temático de contenido para identificar patrones en las disposiciones sobre uso de IA en las distintas fases del proceso de investigación y en la incorporación de principios de integridad, transparencia y responsabilidad (Chen et al., 2024; Limongi, 2024). Finalmente, se compararon los marcos según su atención a las ciencias sociales y las humanidades, en diálogo con aportes recientes sobre ética de la IA en investigación social y uso responsable de IA generativa, con el fin de identificar convergencias, divergencias y vacíos que orienten recomendaciones y futuras líneas de investigación (Jeon et al., 2024; Knöchel et al., 2025).

Resultados

Los resultados de la revisión se organizaron en torno a las tres preguntas de investigación. Respecto de la primera, sobre qué marcos regulatorios y lineamientos éticos sobre el uso de IA en la investigación científica se han desarrollado en los últimos años, la evidencia muestra que la literatura reciente se ha concentrado en describir riesgos y necesidades de orientación más que en sistematizar comparativamente dichos marcos. Chen et al. (2024) analizan, mediante un estudio analítico-descriptivo basado en casos y literatura, el impacto de la IA en la integridad de la investigación y documentan cómo su uso incrementa eficiencia y precisión, pero también facilita nuevas formas de mala conducta, como fabricación de datos y plagio asistido por algoritmos, lo que amenaza la confiabilidad del registro científico. Plantean la urgencia de guías integrales de integridad específicas para IA —con

protocolos claros para análisis de datos y publicación, formación en ética de IA y mecanismos de revisión robustos—, pero no ofrecen una cartografía sistemática por niveles ni un tratamiento diferenciado de disciplinas como las ciencias sociales y las humanidades. De forma convergente, Limongi (2024) analiza, desde una perspectiva teórico-documental, la evolución del uso de IA en investigación y los retos de integridad asociados, mostrando que la IA se ha vuelto indispensable, pero mantiene preocupaciones persistentes sobre sesgos, transparencia, responsabilidad y autoría, lo que exige gobernanza ética reforzada. El estudio destaca la necesidad de algoritmos transparentes y auditables, educación continua y marcos de gobernanza sólidos, a la vez que evidencia la fragmentación de los lineamientos y la ausencia de análisis comparativos por nivel y de atención específica a las necesidades de las ciencias sociales y las humanidades.

En relación con la segunda interrogante, acerca del alcance y contenido principal de los marcos y lineamientos en las distintas fases del proceso de investigación, los resultados indican que las contribuciones tienden a centrarse en principios generales de ética de la IA, con orientaciones desiguales según la fase del ciclo científico. Resnik y Hosseini (2025) examinan, desde un enfoque conceptual y normativo, beneficios y problemas éticos del uso de IA en la investigación, revisando literatura y documentos y proponiendo recomendaciones para un uso responsable. Identifican que la IA plantea retos específicos a la objetividad, reproducibilidad, transparencia, responsabilidad y confianza, y sostienen que no es necesario reemplazar las normas éticas básicas, sino desarrollar guías nuevas y más detalladas sobre cuándo y cómo utilizar estas herramientas. Proponen que los investigadores asuman la responsabilidad de identificar y mitigar sesgos, que se definan protocolos para el uso de IA en el análisis y la redacción y que se exija transparencia en su declaración, pero reconocen que sus recomendaciones siguen siendo generales y que persiste un déficit de marcos que regulen de forma diferenciada cada etapa del proceso de investigación, especialmente fuera del ámbito biomédico.

Chetwynd (2024), enfocada en la escritura científica, revisa literatura y políticas editoriales sobre IA generativa, mostrando que un número creciente de revistas exige declarar el uso de IA, reafirma la responsabilidad exclusiva de los autores humanos y rechaza la coautoría de sistemas de IA. Si bien ello supone un avance en la fase de redacción y publicación, el estudio revela que la mayoría de las directrices no aborda con similar detalle el diseño de estudios ni el análisis automatizado de datos, evidenciando la falta de marcos integrales que articulen normas específicas para todas las fases del ciclo de investigación.

Sobre la tercera pregunta, en torno a la incorporación de disposiciones explícitas y diferenciadas para las ciencias sociales y las humanidades, los hallazgos confirman que la atención disciplinar sigue siendo limitada y fragmentaria. Jeon et al. (2024) abordan la ética de la IA generativa en investigación en ciencias sociales mediante un estudio cualitativo con entrevistas a investigadores implicados en proyectos comunitarios asistidos por IA. Sus resultados muestran una marcada desalineación entre las recomendaciones éticas generales sobre IA y las necesidades específicas de la investigación social, especialmente en lo relativo a la relación con comunidades, la sensibilidad de los datos y la interpretación contextualizada de hallazgos.

Concluyen que la ética de la IA en investigación social debe construirse de forma situada y participativa, mediante marcos de deliberación comunitaria que consideren dinámicas locales e impactos sociales, lo que sugiere que los marcos regulatorios formales no recogen adecuadamente las particularidades de las ciencias sociales y las humanidades y obligan a los investigadores a generar soluciones *ad hoc.* Por su parte, Knöchel et al. (2025) formulan, a partir de un proceso de consenso y análisis conceptual, principios nucleares para el uso responsable de IA generativa en investigación, como transparencia, control humano significativo, responsabilidad, protección de poblaciones vulnerables y evaluación de impactos, presentados como marco transversal aplicable a distintas disciplinas. Aunque reconocen la importancia del contexto disciplinar y de los efectos sociales y comunitarios, su propuesta permanece en un nivel general y no desarrolla lineamientos específicos para áreas como las ciencias sociales y las humanidades, lo que los propios autores identifican como una agenda futura. Esto refuerza el vacío evidenciado por la revisión: la ausencia de marcos normativos y lineamientos éticos que integren de forma explícita y sistemática las particularidades de las ciencias sociales y las humanidades en los distintos niveles de gobernanza (Tabla 1).

Tabla 1

Principales hallazgos por pregunta de investigación y estudios revisados

Pregunta de investigación	Estudio de referencia	Enfoque / metodología	Resultados centrales	Vacío temático identificado
1. Marcos regulatorios y lineamientos éticos vigentes sobre el uso de la IA en la investigación científica.	Chen et al. (2024).	Análisis analítico y descriptivo de casos y literatura sobre IA e integridad científica.	La IA mejora eficiencia y precisión, pero facilita nuevas formas de conducta no deseable (fabricación de datos, plagio asistido), amenazando la confiabilidad del registro científico; se propone desarrollar guías integrales específicas para IA (protocolos para análisis de datos, publicación, formación en ética y revisión rigurosa).	No se ofrece una cartografía sistemática de marcos por nivel (internacional, nacional, institucional) ni se trata de forma diferenciada las ciencias sociales y las humanidades.
	Limongi (2024).	Análisis teórico y documental sobre uso de IA, integridad y ética en investigación.	Muestra que la IA se ha vuelto indispensable, pero mantiene preocupaciones sobre sesgos, transparencia, responsabilidad y autoría; enfatiza la necesidad de algoritmos transparentes, educación continua y marcos de gobernanza sólidos.	Evidencia fragmentación de lineamientos y ausencia de análisis comparativos por nivel y de atención específica a necesidades de ciencias sociales y humanidades.

Tabla 1

Principales hallazgos por pregunta de investigación y estudios revisados

Pregunta de investigación	Estudio de referencia	Enfoque / metodología	Resultados centrales	Vacío temático identificado
2. Alcance y contenido de los marcos según fases del proceso de investigación.	Resnik & Hosseini (2025).	Enfoque conceptual y normativo; revisión de literatura y documentos, con recomendaciones de uso responsable	Identifican impactos de la IA en objetividad, reproducibilidad, transparencia, responsabilidad y confianza; proponen declarar el uso de IA, mitigar sesgos y mantener responsabilidad humana, sin modificar principios éticos básicos.	Las recomendaciones siguen siendo generales; falta regulación diferenciada por etapas del proceso de investigación, especialmente fuera del ámbito biomédico.
	Chetwynd (2024).	Revisión de literatura y políticas editoriales sobre IA generativa en escritura científica.	Muestra que muchas publicaciones exigen declarar uso de IA, reafirman responsabilidad exclusiva de autores humanos y rechazan la coautoría de sistemas de IA.	Las directrices se concentran en redacción y publicación, sin abordar con similar detalle el diseño de estudios ni el análisis automatizado de datos.

Tabla 1

Principales hallazgos por pregunta de investigación y estudios revisados

Pregunta de investigación	Estudio de referencia	Enfoque / metodología	Resultados centrales	Vacío temático identificado
	Jeon et al. (2024).	Estudio cualitativo mediante entrevistas a investigadores en proyectos comunitarios con IA.	Revela desalineación entre recomendaciones éticas generales y necesidades específicas de la investigación social (relación con comunidades, datos sensibles, interpretación contextualizada); propone marcos de deliberación comunitaria situados.	Los marcos regulatorios formales no recogen adecuadamente particularidades de las ciencias sociales y humanidades, obligando a soluciones ad hoc.
3. Disposiciones específicas para ciencias sociales y humanidades.	Knöchel et al. (2025).	Proceso de consenso y análisis conceptual para formular principios nucleares.	Establecen principios transversales (transparencia, control humano significativo, responsabilidad, protección de poblaciones vulnerables, evaluación de impactos) como marco general para el uso responsable de IA generativa.	La propuesta permanece en un nivel general; no desarrolla lineamientos específicos para ciencias sociales y humanidades, identificándose esta adaptación disciplinar como agenda futura.

Nota. Elaboración propia a partir de Chen et al. (2024), Chetwynd (2024), Jeon et al. (2024), Knöchel et al. (2025), Limongi (2024) y, Resnik y Hosseini (2025).

Discusión y conclusiones

La discusión se articula en torno al objetivo de identificar y caracterizar marcos regulatorios y lineamientos éticos sobre el uso de inteligencia artificial (IA) en la investigación científica entre 2019 y 2025, atendiendo a su alcance multinivel y a la inclusión de disposiciones específicas para las ciencias sociales y las humanidades. Los resultados confirman un amplio consenso sobre la necesidad de reforzar la orientación ética y normativa, pero también la ausencia de un mapeo sistemático y comparativo de dichos marcos: mientras Chen et al.

(2024) y Limongi (2024) describen nuevos riesgos para la integridad, sesgos, opacidad y problemas de autoría, ponen el acento en la urgencia de guías y gobernanza robustas más que en una cartografía estructurada de instrumentos existentes. En términos de alcance y contenido, los hallazgos convergen con lo planteado por Resnik y Hosseini (2025) al mostrar que la IA impacta objetividad, reproducibilidad, transparencia, responsabilidad y confianza, y exige declaración de uso, mitigación de sesgos y rendición de cuentas humana; sin embargo, evidencian que la regulación se concentra de forma desproporcionada en la fase de redacción y publicación, tal como reflejan las políticas editoriales analizadas por Chetwynd (2024), dejando relativamente desatendidas etapas como el diseño de estudios y el análisis automatizado de datos. Un aporte central de la revisión es mostrar que la brecha entre recomendaciones generales y necesidades situadas de las ciencias sociales, descrita por Jeon et al. (2024), tiene una base normativa: los marcos rara vez incluyen disposiciones explícitas y diferenciadas para las ciencias sociales y las humanidades, y principios transversales como los propuestos por Knöchel et al. (2025) aún no se han traducido en guías disciplinares específicas.

Asimismo, el estudio evidencia que instrumentos como la Recomendación de la UNESCO sobre ética de la IA y las directrices europeas para el uso responsable de IA generativa inspiran marcos nacionales e institucionales, pero su traducción a reglas operativas concretas —especialmente en las ciencias sociales y las humanidades— es desigual y, a menudo, se limita a reiterar principios generales (UNESCO, 2021; European Commission & European Research Area Forum, 2024). Entre las principales limitaciones de la revisión destacan la restricción lingüística a documentos en inglés y español, la heterogeneidad de las fuentes (artículos científicos, recomendaciones internacionales, guías nacionales, políticas institucionales) y la posible subestimación de enfoques sensibles a las ciencias sociales y las humanidades que no utilizan de forma explícita esa terminología, todo ello en un contexto regulatorio altamente dinámico (Page et al., 2021). De estas constataciones se derivan líneas claras para investigaciones futuras: estudios empíricos sobre la implementación real de marcos en instituciones específicas, análisis comparativos entre países acerca de la "traducción" de recomendaciones internacionales en políticas internas, y el desarrollo de lineamientos específicos para las ciencias sociales y las humanidades que integren principios generales con prácticas disciplinares, idealmente en el marco de revisiones "vivas" que actualicen periódicamente el panorama normativo (Jeon et al., 2024; Knöchel et al., 2025; UNESCO, 2021).

En conjunto, los hallazgos muestran que, aunque se ha avanzado en la identificación de riesgos éticos, en la formulación de principios de integridad y en la propuesta de orientaciones generales para el uso responsable de la IA en la investigación científica, persisten una fuerte fragmentación y heterogeneidad de marcos y lineamientos, una sobrefocalización en la fase de redacción y publicación y una atención escasa, dispersa y poco sistemática a las particularidades de las ciencias sociales y las humanidades. La revisión sistemática cumple el objetivo planteado al ofrecer un mapa crítico y estructurado de los marcos existentes, distinguiendo su alcance y nivel, y señalando con claridad los vacíos que subsisten. Ello subraya la necesidad de avanzar hacia marcos regulatorios y lineamientos éticos más coherentes, operativos y sensibles al contexto disciplinar, y de promover estudios empíricos y revisiones actualizables que acompañen la rápida evolución tecnológica (véase Tabla 2). Este trabajo busca, precisamente, servir como base para el fortalecimiento de una gobernanza de la inteligencia artificial en la investigación que sea, a la vez, técnicamente informada, éticamente robusta, jurídicamente clara y atenta a la diversidad de campos del conocimiento.

Tabla 2

Síntesis de la discusión y proyección normativa sobre el uso de la IA en investigación científica

Eje analítico	Coincidencias con la literatura	Vacíos identificados	Implicaciones para la gobernanza de la IA
Necesidad de regulación y orientación ética.	Se confirma un consenso amplio sobre la urgencia de reforzar la orientación ética y normativa ante nuevos riesgos de integridad, sesgos, opacidad y problemas de autoría (Chen et al., 2024; Limongi, 2024; Resnik & Hosseini, 2025).	No existe aún un mapeo sistemático y comparativo de marcos por nivel (internacional, nacional, institucional), ni una visión de conjunto de los instrumentos vigentes.	Es imperativo pasar de diagnósticos generales a ejercicios de cartografía normativa que permitan identificar solapamientos, lagunas y contradicciones entre marcos.
Alcance de los marcos en el ciclo de investigación.	La literatura destaca que la IA afecta objetividad, reproducibilidad, transparencia y responsabilidad, y exige declaración de uso, mitigación de sesgos y rendición de cuentas humana (Resnik & Hosseini, 2025; Chetwynd, 2024).	La regulación se concentra en la fase de redacción y publicación, mientras que diseño de estudios y análisis automatizado de datos permanecen poco normados.	Se requieren marcos integrales que cubran todas las fases del proceso de investigación, no solo la producción y publicación de manuscritos.

Tabla 2

Síntesis de la discusión y proyección normativa sobre el uso de la IA en investigación científica

Eje analítico	Coincidencias con la literatura	Vacíos identificados	Implicaciones para la gobernanza de la IA
Atención a ciencias sociales y humanidades.	Estudios cualitativos muestran una brecha entre recomendaciones generales y necesidades situadas de la investigación social (Jeon et al., 2024) y proponen principios transversales para el uso responsable de IA generativa (Knöchel et al., 2025).	Los marcos rara vez incluyen disposiciones explícitas y diferenciadas para las ciencias sociales y humanidades; los principios generales aún no se han traducido en guías disciplinares específicas.	Es necesario desarrollar lineamientos específicos para las ciencias sociales y humanidades que integren principios generales con prácticas y contextos disciplinarios (por ejemplo, trabajo comunitario, datos sensibles, métodos cualitativos).
Influencia de marcos internacionales.	La Recomendación de la UNESCO y las directrices europeas inspiran marcos nacionales e institucionales, aportando principios de derechos humanos, justicia, transparencia e integridad en investigación (UNESCO, 2021; European Commission & European Research Area Forum, 2024).	La "traducción" de estos principios en reglas operativas concretas es desigual; en muchos casos se limita a replicar principios sin mecanismos claros de implementación y supervisión, especialmente en las ciencias sociales y humanidades.	Es preciso estudiar y fortalecer los procesos de adaptación local de estos marcos, promoviendo coherencia multinivel y criterios operativos claros para su implementación.
Proyección y agenda de investigación.	Se señalan líneas futuras: estudios empíricos sobre implementación institucional, análisis comparados entre países y revisiones "vivas" que actualicen el mapa regulatorio en un entorno dinámico (Page et al., 2021; Jeon et al., 2024; Knöchel et al., 2025).	La evidencia disponible es aún incipiente respecto al desempeño real de los marcos y su impacto diferenciado en disciplinas y contextos.	Avanzar hacia marcos regulatorios más coherentes, operativos y sensibles al contexto disciplinar permitirá una gobernanza de la IA que sea técnica y éticamente robusta, jurídicamente clara y atenta a la diversidad de campos del conocimiento.

Nota. Elaboración propia a partir de Chen et al. (2024), Chetwynd (2024), European Commission y European Research Area Forum (2024), Jeon et al. (2024), Knöchel et al. (2025), Limongi (2024), Page et al. (2021), Resnik y Hosseini (2025) y UNESCO (2021).

Referencias

Chetwynd, E. (2024). Ethical use of artificial intelligence for scientific writing: Current trends. *Journal of Human Lactation*, 40(2), 211–215. https://doi.org/10.1177/08903344241235160

Chen, Z., Chen, C., Yang, G., He, X., Chi, X., Zeng, Z., & Chen, X. (2024). Research integrity in the era of artificial intelligence: Challenges and responses. *Medicine*, 103(27), e38811. https://doi.org/10.1097/MD.0000000000038811

European Commission, & European Research Area Forum. (2024, March 20). *Guidelines for the responsible use of generative AI in research* [Living guidelines]. European Commission. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/guidelines-responsible-use-generative-ai-research-developed-european-research-area-forum-2024-03-20_en

Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Interamericana Editores.

Jeon, J., Kim, L., & Park, J. (2024). The ethics of generative AI in social science research: A qualitative approach for community-based AI research ethics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4784555

Knöchel, T. D., Schweizer, K. J., Acar, O. A., Akil, A. M., Al-Hoorie, A. H., Buehler, F., Elsherif, M. M., Giannini, A., Heyselaar, E., Hosseini, M., Ilangovan, V., Kovacs, M., Lin, Z., Liu, M., Peeters, A., van Ravenzwaaij, D., Vranka, M. A., Yamada, Y., Yang, Y. F., & Aczel, B. (2025). Core principles of responsible generative AI usage in research. *AI and Ethics*, *5*(6), 6371–6377. https://doi.org/10.1007/s43681-025-00768-8

Limongi, R. (2024). The use of artificial intelligence in scientific research with integrity and ethics. *Future Studies Research Journal: Trends and Strategies*, *16*(1), e845. https://doi.org/10.24023/FutureJournal/2175-5825/2024.v16i1.845

Medina-Romero, M. Á. (2025). *Introducción a la inteligencia artificial en la investigación social*. Editorial SciELa. https://doi.org/10.62131/978-9942-573-14-8

Medina-Romero, M. Á., & Torres Chávez, T. H. (2025). Regulación de la Inteligencia Artificial: Desafíos para los Derechos Humanos en México. *RIDE*. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 15(30). https://doi.org/10.23913/ride.v15i30.2291

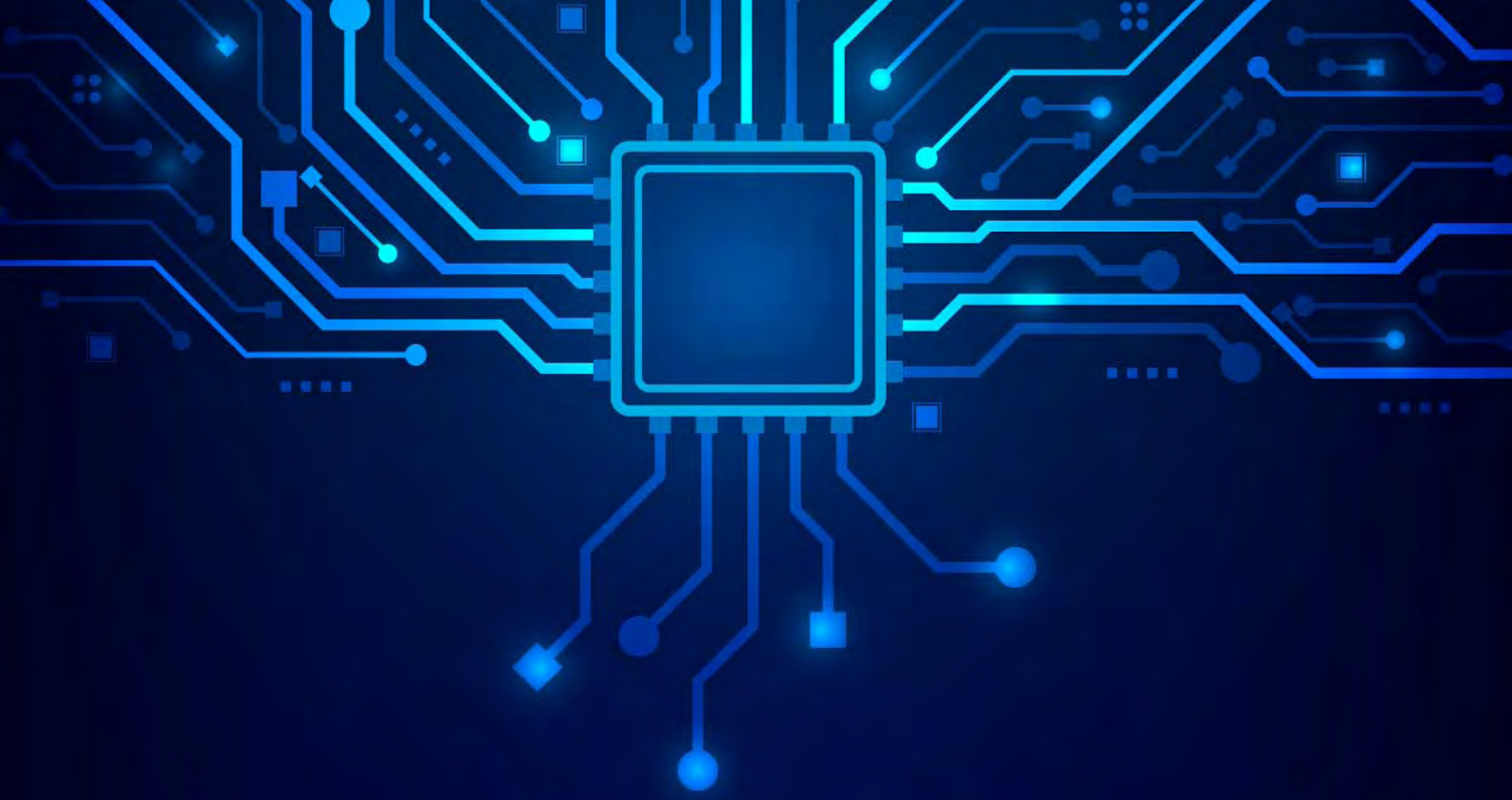
Medina-Romero, M. Á., Hurtado Tiza, D. R., Muñoz Murillo, J. P., Ochoa Cervantez, D. O., & Izundegui Ordóñez, G. (2023). *Método mixto de investigación: Cuantitativo y cualitativo*. Instituto Universitario de Innovación, Ciencia y Tecnología IMUDI Perú. https://doi.org/10.35622/inudi.b.105

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*, *62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Resnik, D. B., & Hosseini, M. (2025). The ethics of using artificial intelligence in scientific research: New guidance needed for a new tool. *AI and Ethics*, *5*(2), 1499–1521. https://doi.org/10.1007/s43681-024-00493-8

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial* (SHS/BIO/REC-AIETHICS/2021). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380455



23.

Inteligencia artificial en el proceso de investigación científica: usos, experiencias y limitaciones

Emma Patricia Mercado-López
Universidad Autónoma de Querétaro, México
ORCID: 0000-0003-0251-6783

Clara Rosalva Mercado-López
Universidad Anáhuac Cancún, México
ORCID: 0009-0004-3723-9012

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.23

23.

Inteligencia artificial en el proceso de investigación científica: usos, experiencias y limitaciones

Introducción

En menos de una década, la inteligencia artificial (IA) ha pasado de ser objeto de estudio especializado a convertirse en herramienta de uso esencial en el ámbito académico. Plataformas conversacionales como *Perplexity*, *Consensus* o *Gemini*, así como asistentes integrados en gestores bibliográficos, hojas de cálculo y programas de análisis cualitativo, ya forman parte del ecosistema del trabajo académico. Esta integración acelerada reconfigura el ciclo de la investigación científica, desde la formulación del problema hasta la comunicación de resultados, al tiempo que plantea interrogantes epistemológicos, éticos y pedagógicos todavía abiertos.

La literatura reciente muestra que los investigadores utilizan sistemas de IA para tareas como la búsqueda y síntesis de información, la depuración y el análisis de datos, la redacción y revisión de manuscritos y la difusión de resultados en diversos formatos. Al mismo tiempo, proliferan advertencias sobre sesgos algorítmicos, opacidad de modelos, vulneración de la autoría académica y riesgo de dependencia tecnológica. Las experiencias reportadas en la investigación indican que, si bien la IA puede aumentar la productividad y facilitar procesos de alfabetización académica, también puede reforzar inequidades de acceso.

Este capítulo se propone explorar críticamente la integración de herramientas de IA en la práctica de la investigación científica. El texto se organizó siguiendo las fases del ciclo de investigación: a) búsqueda y gestión de información con *Consensus*, *Perplexity* y *Gemini*, b) análisis de datos cuantitativos y cualitativos, c) redacción y comunicación científica, y d) construcción de redes referenciales y análisis de contenido. En cada sección se describen y comparan casos de uso, se examinan beneficios y limitaciones y se enfatiza que la IA debe entenderse como herramienta complementaria, no sustitutiva del juicio crítico ni de la responsabilidad epistemológica de quien investiga.

Desarrollo

La inteligencia artificial en la búsqueda y gestión de información científica

Históricamente, la búsqueda de información científica se apoyaba en catálogos bibliotecarios y bases de datos especializadas. La irrupción de motores de búsqueda académicos como *Google Scholar* simplificó este proceso, pero mantuvo la carga de filtrado, lectura y síntesis en la persona investigadora. Sin embargo, con la aparición de asistentes conversacionales con acceso a literatura científica como *Perplexity*, *Gemin*i o *Consensus se* introdujo un cambio cualitativo y cuantitativo de busqueda y gestión de información, debido a que estos sistemas localizan referencias, formulan preguntas en lenguaje natural, obtienen síntesis de resultados, resúmenes estructurados de artículos, extraen variables clave, proponen marcos teóricos y generan tablas de comparación a partir de múltiples fuentes, con lo que reducen significativamente el tiempo de cribado inicial (Marcos-Pablos & García-Peñalvo, 2018; Sabarmathi & Kalamani, 2025).

Consensus, Perplexity y *Gemini* son considerados como herramientas de IA destacadas para busqueda y gestión de información científica, debido a que combinan modelos de lenguaje, recuperación de información y funciones de gestor de referencias. Aunque las tres comparten el objetivo de hacer más eficiente el trabajo documental, difieren en su foco, arquitectura y nivel de especialización para la busqueda y gestión de informacióninvestigación científica (Apata et al., 2025; Faix, 2025; Tjosvold et al., 2025).

Consensus como herramienta de busquedad y gestión de información

Consensus se define explícitamente como una *AI academic search engine,* orientada a literatura revisada por pares. Su núcleo se apoya en una base de más de 200 millones de artículos y *preprints* científicos, sobre la cual aplica búsquedas semánticas y resúmenes automáticos. La herramienta permite formular preguntas en lenguaje natural y devuelve respuestas sintéticas respaldadas por citas visibles. Además, posee un indicador de acuerdo llamado *Consensus Meter*, que muestra el grado de convergencia de la evidencia. La función de busqueda genera en pocos minutos reportes estructurados con estudios clave, posturas divergentes y posibles vacíos de investigación, lo que la hace especialmente útil para revisiones de literatura y mapas de evidencia. En términos de flujo de trabajo académico, *Consensus* se posiciona como un sistema integral de búsqueda y organización de información.

Algunas investigaciones analizaron la herramienta de *Consensus* como buscador académico con IA para bibliotecas universitarias. Sus resultados mostraron los siguiente beneficios: 1) Agiliza la revisión de literatura al responder en lenguaje natural y entregar rápidamente una síntesis de la evidencia, evitando búsquedas manuales extensas. Este ahorro de tiempo es especialmente útil en la investigación ante la alta demanda de información; 2) La facilidad de uso de la interfaz, diseñada con referencias codificadas por color y filtros avanzados se vuelve atractiva para estudiantes de investagación que comienzan a familiarizarse con la literatura científica y reduce la complejidad inicial del proceso de búsqueda; 3) Actúa como un asistente automatizado para bibliotecarios, al apoyar la orientación de búsquedas iniciales, la identificación de estudios clave y la complementación de las bases de datos suscritas por la institución (Apata et al., 2025; Faix, 2025).

No obstante, las investigaciones también han demostrado algunas limitaciones con respecto a la cobertura desigual entre disciplinas: el rendimiento de *Consensus* es mucho mejor en ciencias experimentales y biomédicas que en ciencias sociales y humanidades, áreas donde la producción no está tan bien representada en *Semantic Scholar*. A ello se suma la dependencia de la calidad del corpus: la herramienta solo puede sintetizar lo que está indexado, por lo que su utilidad se ve restringida por la cantidad y la calidad de los artículos disponibles, que suele ser menor que en los catálogos tradicionales de la biblioteca (Apata et al., 2025; Faix, 2025).

Perplexity como herramienta de busquedad y gestión de información

Perplexity, por su parte, se presenta como un buscador conversacional generalista que ha ganado relevancia en contextos académicos por su énfasis en la citación verificable y la actualización constante de sus fuentes. A diferencia de motores como *Google Scholar*, *Perplexity* sintetiza los resultados en lenguaje natural e incluye al pie la lista de enlaces utilizados para cada respuesta, lo que facilita la comprobación de la información. Estudios de intervención didáctica muestran que la capacitación en *Perplexity* mejora de forma significativa la habilidad de estudiantes de bachillerato para localizar referencias, estructurar trabajos académicos y aplicar normas éticas, incluyendo la prevención del plagio (Nuran et al., 2025). Además, acuerdos recientes con editoriales como *Wiley* permiten que, en instituciones suscritas, el modelo consulte directamente colecciones cerradas y devuelva respuestas basadas en contenidos de alta calidad.

Un estudio exploró el uso de *Perplexity* en un entorno altamente especializado en la búsqueda de literatura para evaluaciones de tecnologías sanitarias que requieren la recopilación artículos revisados por pares, informes institucionales, guías clínicas, datos epidemiológicos y *white papers* que suelen estar pobremente indexados en bases bibliográficas tradicionales. El estudio adoptó un diseño de caso, en el que profesionales de la información realizaron búsquedas iterativas en *Perplexity* mediante *prompts* cuidadosamente diseñados. Posteriormente, se compararon los resultados mediante métodos de búsqueda estándar. Los hallazgos revelaron que *Perplexity* identificó fuentes únicas y difíciles de encontrar, particularmente informes de organismos públicos, documentos de *think tanks* y estadísticas oficiales que no aparecían en las búsquedas convencionales (como *Consensus* y *Gemini*). Además, demostrarón en sus resultados, que el tiempo para localizar esos recursos se redujo de forma sustancial, lo que incrementó la eficiencia del proceso de revisión. No obstante, los autores concluyeron que *Perplexity* debería integrarse como complemento de los flujos de trabajo consolidados, por lo que es necesario verificar manualmente la pertinencia de cada fuente, documentar las estrategias de *prompting* y mantener criterios de evaluación crítica de la evidencia (Tjosvold et al., 2025).

Gemini como herramienta de busquedad y gestión de información

Gemini, el modelo de *Google DeepMind*, ocupa un lugar intermedio entre el asistente generalista y el agente de investigación especializado. La versión *Gemini* con capacidades de *Deep Research*, permite la busqueda y la gestión de información con formulación de consultas y busquedas iterativas. *Gemini*, en su modalidad *Deep Research*, se ha posicionado como un asistente de investigación que combina funciones de buscador y gestor de información científica dentro y fuera del ecosistema *Google*. Según la documentación técnica de *Google, Gemini* está diseñado para abordar tareas complejas de investigación descomponiéndolas en subtareas, explorando fuentes en la *web* y en los contenidos del propio usuario, cuando este lo autoriza, y sintetiza los hallazgos en informes estructurados con citas trazables (Google, 2025).

A diferencia de un buscador tradicional, *Gemini* planifica un plan de busqueda. Este propone una estrategia de búsqueda, la ejecuta mediante iteraciones sucesivas y documenta qué ha encontrado y qué pasos seguirá a continuación. Esta capacidad de planear, ejecutar y revisar permite actuar como un gestor activo del proceso de revisión de literatura, en lugar de limitarse a recuperar documentos. El usuario puede revisar el plan, modificarlo

y pedir que *Gemini* profundice en determinados subtemas, lo que convierte la interacción en un diálogo de co-investigación (Mendonça & Marques, 2024).

En términos de gestión de información, *Gemini* permite subir archivos en distintos formatos (*PDF*, *Word*, hojas de cálculo) e integrar su contenido en el informe final. Los informes generados incluyen secciones temáticas, resúmenes de cada fuente clave y una lista de referencias con enlaces directos, lo que facilita la trazabilidad y la posterior lectura crítica. Además, los resultados pueden exportarse a *Google Docs* o *Slides*, compartirse con equipos y reutilizarse como base para propuestas de proyectos, revisiones narrativas o análisis de contexto

Algunos analistas especialistas en IA describen a *Gemini* como un *cerebro digital* que reúne diversas fuentes, referencias, documentos y notas de reuniones. Asimismo, destacan que las funciones de *Gemini* lo acercan a un verdadero gestor de conocimiento organizacional. Sin embargo, estos especialistas subrayan que *Gemini* no sustituye el juicio humano y añaden que, aunque *Gemini* acelera el trabajo pesado de recopilar y ordenar información, las personas investigadoras siguen siendo responsables de validar afirmaciones clave y de decidir qué evidencia es metodológicamente sólida en cada contexto (Google, 2025; Mendonça & Marques, 2024).

Diferencias entre Consensus, Plerplexity y Gemeni para buscar y gestionar información

Las herramientas *Consesnsus*, *Plerplexity* y *Gemini* comparten un conjunto de capacidades convergentes en cuanto a la busqueda y gestión de información. Primero, emplean modelos de lenguaje para interpretar preguntas en lenguaje natural y realizar búsquedas semánticas más allá de las palabras clave exactas. Segundo, generan resúmenes y explicaciones que reducen el tiempo dedicado a la lectura preliminar. Tercero, proporcionan, en mayor o menor medida, citas y enlaces que permiten rastrear las fuentes originales, requisito indispensable para la busqueda y gestión de información. Estos rasgos las distinguen de buscadores tradicionales centrados sólo en la recuperación de enlaces (Mendonça & Marques, 2024).

No obstante, sus diferencias funcionales y epistemológicas son relevantes para el uso científico. *Consensus* opera sobre un universo controlado de publicaciones académicas y prioriza la transparencia en la síntesis de resultados, lo que la convierte en una herramienta adecuada para formular preguntas empíricas, explorar la existencia de consenso y mapear

debates específicos. Su fortaleza reside en la calidad del *corpus* y la orientación explícita al método científico. *Perplexity*, en cambio, equilibra ciencia y divulgación: indexa tanto artículos como páginas *web*, noticias y enciclopedias, lo que amplía el contexto pero exige mayor vigilancia crítica por parte del usuario. Su ventaja está en la versatilidad: es útil para quienes combinan tareas académicas con búsquedas generales, y su diseño conversacional resulta especialmente accesible para estudiantes en fase de alfabetización científica (Nuran et al., 2025).

Gemini sobresale por su capacidad de orquestar procesos complejos: integra búsqueda *web*, análisis de documentos locales, síntesis extensas y exportación directa a plataformas de trabajo colaborativo. Esto la hace idónea para etapas avanzadas de la investigación, como la preparación de informes técnicos, propuestas de proyectos o revisiones narrativas en las que se requiere combinar literatura científica con marcos normativos, datos institucionales y fuentes diversas. Sin embargo, precisamente por operar sobre la *web* abierta, la calidad y relevancia de sus resultados dependen fuertemente de la habilidad de la persona investigadora para diseñar buenos *prompts* y verificar las fuentes vinculadas (Mendonça & Marques, 2024).

Limitaciones y perspectivas críticas de Consensus, Perplexity y Gemini como herramientas de busquedad y gestión de información

Desde una perspectiva crítica, el uso de *Consensus*, *Perplexity* y *Gemini*, plantea riesgos comunes: sesgos en los *corpus* de entrenamiento (con predominio de literatura anglófona), opacidad parcial de los algoritmos de *ranking* y la tentación de aceptar los resúmenes generados sin acudir a los textos originales. Algunos autores, advierten del *perplexity trap*, que tiende a sobrevalorar documentos lingüísticamente más predecibles, lo que podría distorsionar la percepción de la evidencia. Por ello, es indispensable mantener principios de verificación, triangulación y lectura crítica: los asistentes de IA deben entenderse como atajos para localizar y organizar información, no como sustitutos del juicio epistemológico (Wang et al., 2025).

Otras limitaciones reportadas para la búsqueda y gestión de información están condicionadas por los *corpus* de entrenamiento y los acuerdos de acceso a bases de datos. Estudios recientes muestran que muchos modelos generalistas están sobrerrepresentados por literatura en inglés publicada en revistas de alto impacto, mientras que incluyen de forma limitada producción en español, portugués u otras lenguas periféricas. En consecuencia,

Consensus, *Perplexity* y *Gemini* pueden invisibilizar perspectivas latinoamericanas, africanas o asiáticas, reforzando desigualdades epistémicas ya existentes. Este sesgo de cobertura tiene implicaciones directas para investigadores que trabajan en contextos del Sur global o en campos emergentes donde gran parte de la producción se publica en revistas locales o en informes técnicos no indexados en bases de datos comerciales (Wang et al., 2025).

Herramientas de inteligencia artificial para análisis de datos cuantitativos y cualitativos en la investigación

La fase de análisis de datos se ha convertido en uno de los espacios donde la IA muestra mayor capacidad de transformación dentro de la investigación científica. El crecimiento exponencial de los volúmenes de información ha desbordado los marcos analíticos tradicionales y ha abierto la puerta al uso de algoritmos de aprendizaje automático y modelos generativos que automatizan procesos, descubren patrones y generan hipótesis difíciles de detectar sólo con métodos clásicos (Agate, 2025; Zhang et al., 2025).

En el ámbito de los datos cuantitativos, las técnicas de *machine learning* se emplean para mejorar la calidad de los datos, construir modelos predictivos y acelerar la interpretación de resultados. Algunos trabajos revisaron herramientas de IA aplicados a la depuración de bases de datos sanitarias y mostraron como algunos algoritmos de detección de anomalías permitieron identificar incongruencias, valores atípicos y errores de registro, lo que se traduce en estimaciones estadísticas más robustas (Agate, 2025).

De forma similar, otros estudios integraron interfaces gráficas con modelos de clasificación automática para etiquetar grandes volúmenes de datos de dispersión de rayos X; los cuales disminuyeron el tiempo necesario para anotar datos científicos y mejoraron el entrenamiento posterior de modelos predictivos. Estas experiencias sugieren que la IA automatiza el cálculo y reconfigura el flujo de trabajo experimental, al permitir ciclos iterativos en los que los resultados preliminares alimentan nuevas rondas de adquisición de datos (Chavez, et al. 2025).

En el terreno de los datos cualitativos y mixtos, la IA generativa empieza a utilizarse para asistencia en la codificación, la síntesis temática y el análisis de contenido. Algunas herramientas como *ChatGPT* pueden ayudar a generar códigos iniciales, agrupar segmentos de texto y proponer categorías emergentes a partir de entrevistas o grupos focales. No obstante, los modelos carecen de comprensión contextual y situacional: tienden a priori-

zar expresiones frecuentes y pueden invisibilizar voces minoritarias o matices culturales. Por ello se propone un enfoque de *GenAI-in-the-loop*, en el que el investigador utiliza las sugerencias de la máquina como punto de partida exploratorio, pero mantiene el control sobre las decisiones interpretativas y contrasta constantemente los códigos con los datos originales (Nguyen & Welch 2025).

Más allá de aplicaciones específicas, diversos informes de alcance muestran que la IA está aplicando en múltiples disciplinas como un paradigma general de investigación basada en datos. El reporte AI en *Science* del grupo SAPEA (2024) ofrece un mapa amplio de casos de uso en análisis de datos en investigación, desde el diseño automático de materiales en física y química hasta el descubrimiento de relaciones genéticas en biología o la modelización climática con aprendizaje profundo. En este sentido, algunos autores señalan una clara tendencia, donde la la IA puede desempeñar cuatro roles principales en el proceso científico: 1) Para el diseño de estudios; 2) Para la recolección de datos; 3) Para el análisis; 4) Para la revisión por pares. El rol de analista de datos es precisamente el que hoy se expande con mayor rapidez (Li et al., 2025).

El uso de IA para análisis de datos plantea riesgos de opacidad algorítmica, replicabilidad limitada y generación de resultados que los propios investigadores no comprenden plenamente (Holmes et al., 2019; Messeri & Crockett, 2025). La automatización también puede facilitar prácticas cuestionables, como el *p-hacking* asistido o la selección sesgada de modelos que maximizan el ajuste sin fundamentación teórica. Por ello, múltiples autores recomiendan documentar de manera exhaustiva los pasos seguidos por los algoritmos, mantener registros de versiones de modelos y datasets y promover estándares de transparencia algorítmica comparables a los de los métodos estadísticos tradicionales.

En el análisis de datos, la IA se manifiesta en dos planos. En el ámbito cuantitativo, los algoritmos de aprendizaje automático construyen modelos predictivos, detectar patrones no triviales y manejar bases de datos de gran volumen. En el ámbito cualitativo, los modelos de lenguaje facilitan la codificación inicial de entrevistas, foros o artículos y apoyan el análisis de contenido mediante clasificación temática y extracción de entidades. Estos avances pueden acelerar el trabajo y reducir tareas mecánicas. Sin embargo, algunos autores reflexionan sobre las siguientes interrogantes: ¿qué sucede cuando el modelo introduce categorías que no estaban en el marco teórico original?, ¿hasta qué punto la investigadora delega la interpretación en una *caja negra* entrenada con datos ajenos a su contexto? (Chen et al., 2020).

Integración de herramientas de inteligencia artificial en la redacción y comunicación científica

La integración de herramientas de IA en la redacción y comunicación científica está modificando de manera acelerada la práctica cotidiana de la investigación. Los modelos de lenguaje de gran escala como *Perplexity* o *Gemini* y los asistentes integrados en gestores de referencias, procesadores de texto y plataformas editoriales se han convertido en aliados habituales para planear manuscritos, mejorar el estilo, sintetizar resultados y adaptar contenidos a diferentes audiencias. Sin embargo, su adopción plantea dilemas metodológicos, éticos y de autoría que obligan a repensar qué entendemos por escribir ciencia en la era de la IA.

Estudios recientes mostraron que una proporción creciente de investigadores utilizan IA en actividades de lectura y escritura académica. Asimismo explicaron el impacto de estas herramientas en la escritura científica universitaria, donde se encontró que la mayoría de trabajos revisados presentaron mejoras en la coherencia, la organización discursiva, la riqueza léxica y la calidad argumentativa de los textos científicos cuando el uso de IA se combina con la del investigador (Sanz-Tejeda et al., 2026).

Por otro lado una investigación en Estados Unidos describió que la IA en el ámbito de redacción científica se emplea principalmente por académicos para generar ideas, reorganizar apartados, ofrecer retroalimentación formativa (evaluación de escritos científicos) y como apoyo a los procesos de revisión y autoedición (Arroyo-Machado et al., 2025).

Asimismo, en la práctica, la IA se ha incorporado en tres niveles de la escritura científica. En primer lugar, como herramienta de ideación y planificación: muchos autores recurren a los modelos de lenguaje para explorar preguntas de investigación, generar esquemas de secciones o proponer títulos y resúmenes tentativos. En segundo lugar, como asistente lingüístico para mejorar claridad y corrección: corrección gramatical, reformulación de frases ambiguas, ajuste del registro y adaptación a normas estilísticas específicas (ejemplo normas APA 7) se han convertido en tareas habituales delegadas a la IA. En tercer lugar, algunos investigadores la usan para producir borradores sustanciales de textos que luego son revisados y reescritos extensamente por los autores humanos.

Estas prácticas varían en intensidad, pero comparten la idea de que la IA puede aliviar la carga mecánica de la escritura y permitir concentrarse en la interpretación de resultados y la construcción teórica.

No obstante, la literatura advierte que estos beneficios vienen acompañados de riesgos significativos. por la posibilidad de introducir errores, sesgos o afirmaciones inexactas que, al no ser plenamente comprendidas por el autor, pasen inadvertidas hasta etapas avanzadas del proceso editorial.

Estas inquietudes han motivado respuestas normativas desde la ética de la publicación. El Comité de Ética en Publicaciones (COPE) emitió en 2023 un posicionamiento claro: las herramientas de IA no pueden figurar como autoras porque no cumplen criterios de responsabilidad intelectual, y los autores humanos deben declarar de forma transparente qué herramientas utilizaron y para qué tareas. Además, se subraya que los investigadores siguen siendo plenamente responsables de verificar la exactitud de todo el contenido, incluyendo el generado o editado por IA. Revisiones de alcance sobre la escritura académica con IA señalan que están emergiendo marcos normativos más finos, como taxonomías de tare y modelos de divulgación por niveles que especifican si la IA se usó para ideación, edición lingüística, síntesis de literatura o apoyo en la revisión por pares (Müller & García, 2025).

La IA también se integra en la comunicación ampliada de la ciencia. Los mismos modelos que ayudan a escribir artículos permiten generar resúmenes gráficos, hilos para redes sociales, notas de prensa y materiales divulgativos para audiencias no especializadas. Esto puede democratizar el acceso al conocimiento científico, al facilitar la adaptación de resultados a distintos formatos y niveles de complejidad.

Sin embargo, existe el riesgo de que la simplificación automática distorsione matices metodológicos cruciales o sobredimensione resultados preliminares. Por ello, varias guías recomiendan que cualquier uso de IA en productos de comunicación pública sea revisado por el equipo de investigación, cuidando tanto la fidelidad a los datos como la claridad del mensaje (SAPEA, 2024).

Desde una perspectiva ética y epistemológica, algunos autores proponen entender la IA generativa como una especie de *coautor senior* no humano. El uso intensivo de IA podría considerarse análogo. No obstante, incluso en este marco, se recalca que debe haber siempre una persona responsable que pueda dar cuenta del contenido y asumir las consecuencias de errores o malas prácticas. En contraste, otros trabajos insisten en mantener una separación tajante entre herramientas y autores, para evitar diluir la noción de responsabilidad en un contexto ya complejo de autorías múltiples (Müller & García, 2025).

Integración de herramientas de inteligencia artificial en la construcción y análisis de redes referenciales

La integración de herramientas de IA en la construcción y análisis de redes referenciales y en el análisis de contenido está cambiando de forma profunda la manera en que se cartografía, interpreta y comunica la producción científica. La masificación de publicaciones y la complejidad de los debates hacen cada vez más difícil que una persona investigadora pueda *leerlo todo*. En este contexto, la IA se convierte en un aliado para automatizar tareas bibliométricas, generar mapas de conocimiento y apoyar el análisis temático y discursivo de grandes textos.

En el ámbito de las redes referenciales, las técnicas clásicas de bibliometría y cienciometría —cocitación, coautoría, acoplamiento bibliográfico— se han visto reforzadas por algoritmos de aprendizaje automático y modelos de lenguaje. Algunos trabajos revisaron de manera sistemática aplicaciones de IA en *scientometrics*, *webometrics* y *bibliometrics, para* mostrar que los algoritmos de *clustering*, las redes neuronales y el procesamiento del lenguaje natural permiten automatizar la construcción de redes de publicaciones, autores y palabras clave, así como mejorar la desambiguación de nombres, la detección de comunidades científicas y la predicción de impacto. Estas capacidades son cruciales para detectar escuelas de pensamiento, colaboraciones emergentes y áreas de convergencia entre disciplinas (Saeidnia et al., 2024).

Herramientas de uso extendido como *VOSviewer o Bibliometrix/Biblioshiny* incorporan funciones que, aunque no siempre se etiquetan como IA se apoyan en técnicas de aprendizaje no supervisado para agrupar nodos y visualizar estructuras de red. *VOSviewer* permite construir mapas de coautoría, cocitación y coocurrencia de términos a partir de datos exportados de *Scopus* o *Web of Science*,. Esta utiliza algoritmos de *clustering* que revelan comunidades y temas de investigación de forma intuitiva. Estos sistemas son cada vez más complementados por marcos de IA que pueden traducir consultas en lenguaje natural en cadenas de procesamiento bibliométrico.

La construcción de redes referenciales alimentadas por IA también se relaciona con el desarrollo de *knowledge graphs*. Propuestas como *SCICERO* utilizan modelos de lenguaje y técnicas de extracción de información para identificar entidades (conceptos, métodos, datasets) y relaciones entre ellas en millones de artículos, generando grafos de conocimiento que describen el *paisaje conceptual* de un campo (Hai Duong et al., 2022). Estos grafos permiten

navegar por conexiones no evidentes en lecturas lineales, identificar trabajos puente entre subdisciplinas y apoyar la generación de nuevas hipótesis. De este modo, la IA representa las redes de citas y las redes semánticas que subyacen a la producción científica.

En paralelo, la IA está transformando el análisis de contenido de la literatura y de datos cualitativos. Los modelos *BERT* y sus variantes especializadas en texto científico, como *SciBERT*, se utilizan para clasificación temática, detección de tópicos y extracción de conceptos clave en grandes colecciones de artículos. Por ejemplo, Krishnan et al. (2025) propusieron un modelo híbrido *BERT + TV-Tree* para detectar temas de investigación en miles de artículos y organizar jerárquicamente los tópicos; los temas detectados se utilizaron como etiquetas para mejorar las búsquedas de literatura y para construir mapas temáticos más precisos. Estos enfoques permitieron al equipo de investigadores identificar rápidamente las principales líneas de trabajo y las lagunas en un área determinada.

En el campo de la investigación cualitativa, se introduce el concepto de *conversar con los datos* mediante modelos de lenguaje. En lugar de limitarse al código manual línea por línea, el investigador puede hacer preguntas al modelo sobre un corpus de entrevistas, solicitar ejemplos de citas que respalden ciertas interpretaciones o explorar variaciones en el uso de conceptos (Hayes, 2025) . De forma complementaria, el marco *LLM-Assisted Content Analysis* (LACA) plantea una integración metodológicamente rigurosa de los modelos de lenguaje en la codificación deductiva, donde la IA sugiere códigos y justificaciones textuales, mientras que el analista evalúa, acepta o rechaza cada propuesta (Chew et al., 2023). Estudios empíricos con LACA muestran reducciones significativas en la carga de trabajo y aumentos en la consistencia intercodificador, sin renunciar al control humano sobre las decisiones analíticas.

No obstante, la integración de IA en redes referenciales y análisis de contenido plantea riesgos y desafíos. Saeidnia et al. (2024) advierten que la automatización puede amplificar sesgos de cobertura, por ejemplo, la sobrerrepresentación de literatura en inglés y de revistas de alto impacto— y consolidar métricas de influencia que favorecen a actores ya consolidados. Del lado del análisis de contenido, se subraya que los modelos de lenguaje tienden a privilegiar patrones mayoritarios y pueden invisibilizar voces marginales o interpretaciones alternativas, lo que exige una vigilancia constante sobre los efectos de *normalización* que la IA puede introducir. Además, el uso de datos completos de artículos, entrevistas o documentos institucionales en plataformas propietarias suscita preocupaciones legítimas sobre privacidad, derechos de autor y gobernanza de los datos (Hayes, 2025).

Por ello, la mayoría de autores coincide en que la IA debe entenderse como un amplificador del trabajo analítico, no como un reemplazo del juicio experto. Las redes de cocitación, los grafos de conocimiento y los resúmenes temáticos generados automáticamente son puntos de partida para la reflexión, no conclusiones definitivas. La calidad de los resultados depende, en última instancia, de la capacidad de los equipos para formular buenas preguntas, seleccionar y depurar los datos de entrada, interpretar críticamente las visualizaciones y contrastar cualquier inferencia con el conocimiento sustantivo del campo. En este sentido, la integración de IA en la construcción de redes referenciales y el análisis de contenido exige competencias técnicas y una alfabetización crítica en datos y algoritmos.

Conclusiones

Consensus, Perplexity y *Gemini* ilustran tres caminos complementarios en la integración de la IA al trabajo de búsqueda y gestión de información científica. *Consensus* privilegia la evidencia y la síntesis estructurada de resultados académicos; *Perplexity* destaca por su versatilidad, su orientación a la docencia y su insistencia en la citación verificable; *Gemini* se proyecta como un agente de investigación integral que articula búsqueda, análisis y gestión documental en un mismo entorno. La elección entre uno u otro o su uso combinado debe responder a las necesidades concretas de cada proyecto y al nivel de competencia informacional de quienes investigan. En cualquier caso, el valor real de estas herramientas no reside en la automatización en sí misma, sino en la capacidad de las comunidades científicas para apropiarse críticamente de ellas, adaptarlas a sus contextos y someter sus resultados a escrutinio colectivo.

Las herramientas de inteligencia artificial abren oportunidades sin precedentes para el análisis de datos en la investigación científica. Las cuales, permiten tratar volúmenes masivos de información, automatizar tareas repetitivas, explorar patrones complejos y democratizar el acceso a métodos avanzados. Sin embargo, no constituyen una *solución mágica.* Su uso responsable exige datos de alta calidad, supervisión experta, escrutinio metodológico y marcos éticos que garanticen transparencia, reproducibilidad y equidad. La IA puede y debe entenderse como un colaborador crítico en el análisis de datos, siempre que el control epistemológico permanezca en manos de la comunidad científica.

La integración de herramientas de IA en la redacción y comunicación científica ofrece oportunidades genuinas para mejorar la calidad lingüística, agilizar procesos editoriales y ampliar el alcance de la difusión, en especial para investigadores que escriben en segundas

lenguas o trabajan en entornos con altas cargas docentes. No obstante, estos beneficios solo son compatibles con el ethos científico si se asumen algunos principios: transparencia en el uso de IA, verificación rigurosa de todo el contenido generado, preservación del razonamiento crítico en la construcción de argumentos y reconocimiento explícito de que la responsabilidad última recae siempre en las y los investigadores humanos. Bajo estas condiciones, la IA puede convertirse en una aliada poderosa, pero nunca en un sustituto.

Las herramientas de IA abren también posibilidades inéditas para mapear y analizar la producción científica: permiten construir redes referenciales complejas, modelar tópicos, generar grafos de conocimiento y dialogar con grandes corpus textuales de manera flexible. Sin embargo, su valor para la investigación depende de que se utilicen bajo principios de transparencia, control humano significativo y reflexión epistemológica. Solo así estas tecnologías podrán contribuir a una cartografía más rica, plural y rigurosa del conocimiento, en lugar de reforzar inercias y sesgos ya presentes en los sistemas científicos.

Referencias

Agate, J. (2025). Artificial intelligence methods and approaches to improve data quality in healthcare. *AI in Medicine*, *18*(2), 45–59. https://doi.org/ 10.1016/j.ailsci.2025.100135

Apata, O.E., Kwok, O., & Lee, Y. (2025). The Use of Generative Artificial Intelligence (AI) in Academic Research: A Review of the Consensus App. *Cureus*, 17. https://doi.org/10.7759/cureus.87297

Arroyo-Machado, W., Jinghuan Ma, T., Timothy, P., Shaika Islam, L., Michalegko, E. WelchTipeng C., (2025). Generative AI and academic scientists in US universities. *PLOS ONE*, *20*(8), e0330416. https://doi.org/10.1371/journal.pone.0330416

Chavez, T., Zhao, Z., Jiang, R., Koepp, W., et al. (2025). A machine-learning-driven data labeling pipeline for scientific analysis. *Journal Appl Crystallogr*. https://doi.org/10.1107/S1600576725002328

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chew, B., Tan, J., & Koh, A. (2023). LLM-Assisted Content Analysis (LACA): Integrating large language models into deductive qualitative coding. *International Journal of Qualitative Methods*, 22, 1–15.

Committee on Publication Ethics. (2023). *COPE position statement on authorship and AI tools*. https://publicationethics.org

Faix, A. (2025). Consensus: Using AI to analyze scientific literature. *Library Trends*, *73*(3), 344–354. https://doi.org/10.1353/lib.2025.a961198

Google. (2025). *Create detailed reports with Deep Research Google Workspace Updates*. https://workspace.google.com/blog/

Hai Duong, T., Thong Tran, T., Bac Le, P., Fournier-Viger P. (2022). SCICERO: A deep learning and NLP approach for generating scientific knowledge graphs in the computer science domain. *Knowledge-Based Systems*, 256, 109921. https://doi.org/10.1016/j.knosys.2022.109921

Hayes, A. S. (2025). Conversing with qualitative data: Enhancing qualitative research through large language models. *International Journal of Qualitative Methods*, 24, 1–18. https://doi.org/10.1177/16094069251322346

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Krishnan, K., Easwarakumar, K. S. Research topic detection in scientific articles using a hybrid BERT integrated telescopic vector tree model with emperor penguin enhanced NSGA II optimization. *Sci Rep,* 15, 37201 (2025). https://doi.org/10.1038/s41598-025-21145-9

Li, L.; Xu, K.; Su, R.; Gu, H.; Ma, P. (2025) AI as a catalyst for transforming scientific research: a perspective. *AI Agent*, 1, 8. https://dx.doi.org/10.20517/aiagent.2025.08

Marcos-Pablos, S., & García-Peñalvo, F. J. (2018). Information retrieval methodology for aiding scientific database search. *Soft Computing*, *24*(8), 5551–5560. https://doi.org/10.1007/s00500-018-3568-0

Mendonça, R., & Marques, A. (2024). Scientific research and the digital brain: Managing information and changing the way we think. *Blucher Design Proceedings*, *13*(1), 1134–1145. https://doi.org/10.5151/sigradi2023-432

Messeri, L., & Crockett, M. (2025). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *626*(8002), 37–41. https://doi.org/10.1038/s41586-024-07146-0

Müller, L., & García, R. (2025). The 2025 landscape of generative AI in scholarly writing: Uses, governance and ethical frameworks. *Journal of Scholarly Communication*, 12(4), 1–32.

Nguyen, D. C., & Welch, C. (2025). Generative Artificial Intelligence in Qualitative Data Analysis: Analyzing—Or Just Chatting? *Organizational Research Methods*, *29*(1), 3-39. https://doi.org/10.1177/10944281251377154

Nuran, A. S., Rahmah, A., Hamris, H., Misbahuddin, A. F., & Faisal, A. N. (2025). *Pelatihan integrasi Perplexity untuk penelusuran literatur dan penulisan karya ilmiah di SMAN 5*.

Sabarmathi, K. R., & Kalamani, M. (2025). Scientific retrieval: An effective heuristic-aided multi-scale adaptive transformer network for information retrieval process from scientific publications. *Knowledge and Information Systems*, 67, 8377–8415. https://doi.org/10.1007/s10115-025-02458-7

Saeidnia, H., Badie, K., & Didegah, F. (2024). Unleashing the power of AI: A systematic review of cutting-edge techniques in AI-enhanced scientometrics, webometrics and bibliometrics. *Computer Science*, *18*(3), 101419.

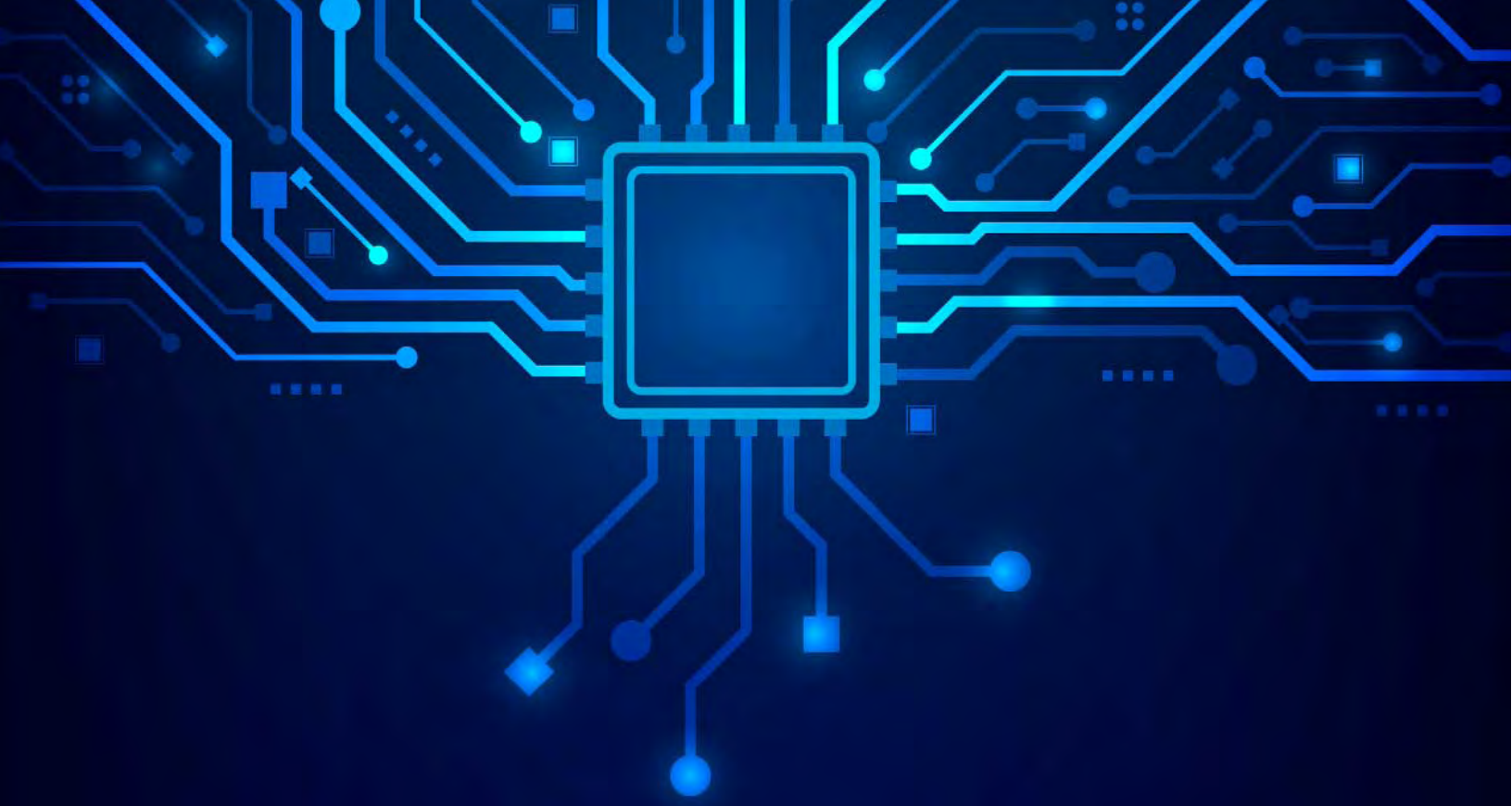
Sanz-Tejeda, A., Domínguez-Oller, J. C., Baldaquí-Escandell, J. M., Gómez-Díaz, R. & García-Rodríguez, A. (2025). The impact of generative AI on academic reading and writing: A systematic review. *Frontiers in Education*, 10, 1711718. https://doi.org/10.3389/feduc.2025.1711718

SAPEA. (2024). *AI in science: Evidence review report*. Science Advice for Policy by European Academies. https://scientificadvice.eu

Tjosvold, L., Reimer, D., & Kung, J. (2025). OP83 Uncovering the hidden evidence: Harnessing Perplexity artificial intelligence to revolutionize grey literature retrieval in health technology assessment. *International Journal of Technology Assessment in Health Care*, *41*(S1), S38. https://doi.org/10.1017/S0266462325101372

Wang, H., Dai, S., Zhao, H., Pang, L., Zhang, X., Wang, G., Dong, Z., Xu, J., & Wen, J.-R. (2025). Perplexity trap: PLM-based retrievers overrate low perplexity documents. *Proceedings of the International Conference on Learning Representations* (*ICLR 2025*). Advance online publication. https://doi.org/10.48550/arXiv.2503.08684

Zhang, X., Jiang, Z., & Wang, Z. (2025). Retrieval-augmented document-level information extraction from scientific articles using large language models. *Data Intelligence*, *7*(1), 1–25. https://doi.org/10.3724/2096-7004.di.2025.0175



24.

La mediación epistémica de la inteligencia artificial generativa en la ciencia contemporánea

Andrés Valencia Sánchez
Universidad Pedagógica del Estado de Sinaloa, México
ORCID: 0009-0002-2482-3802

José Cristóbal Solís Pollorena
Universidad Pedagógica del Estado de Sinaloa, México
ORCID: 0000-0003-0075-7858

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbo12.24

24.

La mediación epistémica de la inteligencia artificial generativa en la ciencia contemporánea

Introducción

El desarrollo reciente de la inteligencia artificial generativa (IAG) ha marcado un punto de inflexión en múltiples ámbitos del quehacer humano, particularmente en la investigación académica y científica. A diferencia de tecnologías previas orientadas a la automatización de tareas específicas, los modelos generativos actuales poseen la capacidad de producir textos, imágenes, códigos, esquemas conceptuales y simulaciones con un nivel de sofisticación que interpela directamente los procesos tradicionales de generación del conocimiento.

En el ámbito académico, plataformas como *ChatGPT*, *Claude* o *Gemini* son utilizadas cotidianamente para la delimitación de problemas de investigación, la redacción de marcos teóricos, el análisis preliminar de datos y la mejora del estilo académico. De manera paralela, herramientas visuales como *Stable Diffusion*, *Midjourney* o *DALL·E* han ampliado las posibilidades de representación, modelización y divulgación científica. No obstante, esta adopción acelerada ha ocurrido, en muchos casos, sin una reflexión epistemológica profunda sobre el papel que dichas plataformas desempeñan en el proceso investigativo.

El problema central que aborda este artículo radica en la tendencia a concebir la IAG exclusivamente como una herramienta instrumental, neutral y subordinada al investigador, invisibilizando su capacidad de influir en la formulación de preguntas, en la organización del conocimiento y en los criterios de validez científica. Desde esta perspectiva, resulta pertinente preguntarse: ¿de qué manera las plataformas de IAG están transformando la lógica interna de la investigación académica? ¿Cómo inciden en la relación entre el sujeto investigador, el método y el objeto de estudio?

Ante esto, el objetivo del presente trabajo es explicar el papel de las plataformas de IAG como mediadores epistémicos en la investigación académica y científica, identificando

sus aplicaciones más relevantes, sus potencialidades y sus límites. La relevancia del estudio estriba en la necesidad de construir marcos teóricos que orienten un uso crítico, responsable y formativo de la IA, evitando tanto su rechazo acrítico como su adopción irreflexiva.

Marco teórico-conceptual

Inteligencia artificial generativa: delimitación conceptual

La IAG puede definirse como el conjunto de sistemas computacionales capaces de producir contenidos nuevos entre ellos los textuales, visuales, sonoros o multimodales a partir del aprendizaje de grandes volúmenes de datos. Estos sistemas, se apoyan fundamentalmente en modelos fundacionales de gran escala, los cuales son entrenados mediante arquitecturas de aprendizaje profundo que les permiten identificar patrones complejos y generar salidas coherentes más allá de la simple recuperación de información (Bommasani et al., 2021; Brown et al., 2020).

Desde una perspectiva teórica, la principal diferencia entre la IAG y la inteligencia artificial predictiva o clasificatoria radica en su capacidad de síntesis, producción, y no únicamente en la anticipación de resultados o en la categorización de datos. Mientras que los sistemas predictivos operan bajo lógicas probabilísticas orientadas a la optimización de decisiones específicas, los modelos generativos introducen una dimensión creativa que amplía los límites tradicionales de la automatización, al permitir la construcción de representaciones, narrativas y estructuras conceptuales nuevas (Floridi, 2024).

Esta dimensión creativa no debe entenderse como creatividad en sentido humano, sino como una capacidad de recombinación informacional guiada por estructuras estadísticas, lo cual tiene implicaciones profundas para la investigación científica. En particular, la IAG incide en procesos centrales del quehacer académico, como la formulación de hipótesis, la articulación de marcos teóricos y la interpretación preliminar de datos, actuando como un mediador epistémico entre el investigador y el objeto de estudio (Latour, 2008; Hutchins, 1995).

Desde la epistemología social del conocimiento, estos sistemas pueden ser comprendidos como dispositivos que condensan prácticas, consensos y sesgos provenientes de comunidades científicas amplias, influyendo en la manera en que el conocimiento es producido y validado (Longino, 1990). En consecuencia, la relevancia de la IAG para la in-

vestigación científica no reside únicamente en su eficiencia operativa, sino en su capacidad para reconfigurar las dinámicas cognitivas y metodológicas que sostienen la producción del conocimiento académico.

La inteligencia artificial como mediación epistémica

Desde una perspectiva epistemológica contemporánea, la relación entre tecnología y conocimiento ha sido problematizada por enfoques que cuestionan la concepción clásica de la técnica como un mero instrumento neutral al servicio del sujeto cognoscente. En particular, las corrientes post-positivistas y post-humanistas han subrayado que los artefactos tecnológicos participan activamente en la configuración de las condiciones de posibilidad del conocer, al intervenir en los modos de percepción, interpretación y validación de la realidad (Verbeek, 2011; Simondon, 2008).

En este marco, la tecnología puede ser comprendida como una forma de mediación epistémica, en tanto no solo amplía las capacidades humanas, sino que reconfigura las prácticas cognitivas y los horizontes de sentido desde los cuales se produce el conocimiento. La mediación tecnológica, lejos de ser transparente, introduce orientaciones, restricciones y sesgos que influyen en la construcción de los objetos de investigación y en la articulación de los discursos científicos (Feenberg, 2002).

Aplicada al contexto de la IAG, esta perspectiva permite sostener que plataformas como *ChatGPT* o *Claude*, no operan únicamente como dispositivos de apoyo funcional, sino como infraestructuras cognitivas activas que participan en la estructuración del razonamiento académico. La generación de conceptos, la reorganización argumentativa, la jerarquización de fuentes y la reformulación de marcos teóricos constituyen prácticas en las que la IA interviene como un co-configurador del proceso cognoscitivo, influyendo de manera indirecta pero significativa en las decisiones epistemológicas del investigador (Zuboff, 2020; Kitchin, 2014).

Desde la epistemología social del conocimiento, esta mediación adquiere una dimensión estructuralmente colectiva. Los modelos de IAG son entrenados sobre vastos repositorios textuales que condensan tradiciones disciplinares, consensos científicos dominantes y patrones discursivos históricamente situados. En este sentido, la IA no solo reproduce información, sino que incorpora formas socialmente estabilizadas de razonamiento científico, actuando como un artefacto epistémico que cristaliza prácticas cognitivas agrupadas (Knorr Cetina, 2005).

Asimismo, desde una perspectiva post-humanista del conocimiento, resulta pertinente cuestionar la centralidad exclusiva del sujeto humano en la producción científica. Autores como Hayles (2017), han señalado que los sistemas inteligentes introducen nuevas formas de agencia distribuida, en las que la producción del conocimiento emerge de ensamblajes híbridos entre humanos, algoritmos y entornos digitales. Bajo esta óptica, la IAG puede ser entendida como un agente no humano con capacidad de incidencia epistemológica, cuya participación redefine los límites entre autoría, interpretación y validación científica.

En consecuencia, el análisis de la IAG en la investigación académica exige superar enfoques meramente instrumentales y adoptar marcos teóricos que permitan comprender su papel como mediador epistémico no humano. Este desplazamiento conceptual resulta fundamental para abordar críticamente los desafíos metodológicos, éticos y ontológicos que plantea la integración de la inteligencia artificial en los procesos contemporáneos de producción del conocimiento científico.

DESARROLLO

Modelos de lenguaje de gran escala

Los modelos de lenguaje de gran escala (*Large Language Models*, LLM) constituyen uno de los desarrollos más relevantes de la IA contemporánea aplicada al ámbito académico y científico. Plataformas como *ChatGPT*, *Claude*, *Gemini* y *DeepSeek* se basan en arquitecturas de aprendizaje profundo, principalmente modelos de tipo *transformer*, entrenadas con grandes volúmenes de datos textuales, lo que les permite procesar, comprender y generar lenguaje natural con altos niveles de coherencia semántica y contextual (Russell & Norvig, 2004).

En el campo de la investigación académica, estos modelos han sido incorporados progresivamente como herramientas de apoyo cognitivo, especialmente en fases iniciales y de desarrollo del proceso investigativo. Entre sus principales funcionalidades se encuentran la exploración preliminar de temas, la delimitación de objetos de estudio, la formulación de preguntas de investigación, la organización de marcos teóricos, la síntesis de literatura científica y la redacción asistida de textos académicos. Estas funciones, responden a su capacidad para identificar patrones discursivos, estructuras argumentativas y relaciones conceptuales presentes en amplios corpus especializados (García-Peñalvo et al., 2024).

No obstante, el aporte de los modelos de lenguaje de gran escala no se limita a la mejora de la eficiencia operativa del investigador. Desde una perspectiva epistemológica, estos sistemas introducen una transformación cualitativa en los procesos cognitivos asociados a la investigación, al intervenir en la organización del pensamiento, la jerarquización conceptual y la estructuración del discurso académico. Como señala Floridi (2014), la mediación tecnológica avanzada no solo acelera tareas, sino que reconfigura las condiciones mismas de producción del conocimiento, generando nuevas formas de interacción entre el sujeto investigador y la información.

En este sentido, los LLM pueden ser comprendidos como dispositivos de andamiaje cognitivo, en tanto facilitan procesos complejos como la síntesis teórica o la formulación argumentativa, especialmente en contextos de alta densidad informativa. Desde la epistemología de la mediación, esta función implica que el conocimiento ya no se produce exclusivamente a partir de la relación directa entre el investigador y las fuentes, sino mediante una interacción mediada algorítmicamente, en la que el modelo sugiere rutas interpretativas y configuraciones discursivas posibles (Verbeek, 2005).

Entre los principales aportes y bondades de los modelos de lenguaje de gran escala en la investigación académica destacan:

1) La optimización del tiempo cognitivo, al reducir la carga operativa asociada a tareas repetitivas.

2) El acceso ampliado a estructuras conceptuales complejas, lo que favorece la comprensión interdisciplinaria.

3) La facilitación de procesos reflexivos, al permitir contrastar enfoques, reformular argumentos y explorar alternativas teóricas.

Estas ventajas son particularmente relevantes en contextos de investigación formativa y en etapas exploratorias del trabajo científico (UNESCO, 2023). Sin embargo, diversos autores advierten que el uso intensivo de estos modelos plantea desafíos epistemológicos significativos, especialmente en relación con la originalidad, la autonomía intelectual y la autoría académica. Al proponer estructuras argumentativas plausibles y estilos discursivos estandarizados, los LLM pueden influir en la manera en que el investigador conceptualiza su objeto de estudio, lo que exige una actitud crítica y reflexiva frente a sus resultados. Como

subraya Boden (2017): la generación automatizada de contenidos no equivale a creatividad humana, sino a una recombinación sofisticada de patrones previamente aprendidos.

En esta idea, los modelos de lenguaje de gran escala deben ser comprendidos en la investigación académica, no como sustitutos del pensamiento científico, sino como herramientas de mediación cognitiva avanzada, cuyo valor depende del uso crítico, ético y reflexivo que el investigador haga de ellas. Reconocer sus aportes y limitaciones permite integrarlos de manera responsable en la práctica científica, fortaleciendo la calidad argumentativa, sin comprometer la autonomía intelectual ni el rigor epistemológico.

Plataformas visuales y multimodales

Las plataformas generativas visuales y multimodales, como *Stable Diffusion*, *Midjourney* y *DALL·E*, representan una extensión significativa de la IAG hacia el ámbito de la producción de representaciones visuales complejas. Estas herramientas, se basan en modelos de difusión y redes neuronales profundas entrenadas con grandes conjuntos de datos multimodales —texto e imagen—, lo que les permite generar imágenes originales a partir de descripciones lingüísticas (*text-to-image*) o combinar múltiples modalidades de entrada y salida (Russell & Norvig, 2021).

En el campo de la investigación académica y científica, estas plataformas han ampliado de manera sustantiva las posibilidades de visualización, modelización y comunicación del conocimiento. Su uso resulta, particularmente relevante en disciplinas como las ciencias naturales, la ingeniería, la medicina, la arquitectura, las ciencias sociales aplicadas y la educación científica, donde la representación gráfica cumple una función central en la comprensión de fenómenos complejos o abstractos. Como señalan Tufte (2006) y Cairo (2016): la visualización no es un mero recurso ilustrativo, sino una herramienta cognitiva que influye directamente en la forma en que se interpreta y valida la información científica.

Desde una perspectiva funcional, estas plataformas permiten generar modelos conceptuales, simulaciones visuales, esquemas explicativos e imágenes hipotéticas que facilitan la exploración teórica y la divulgación científica. En etapas tempranas de la investigación, pueden emplearse como apoyo para la conceptualización de ideas, la formulación de hipótesis visuales o la anticipación de posibles escenarios, mientras que en fases de comunicación de resultados contribuyen a mejorar la accesibilidad y el impacto social del conocimiento científico (García-Peñalvo, 2024).

Uno de los principales aportes y bondades de las plataformas visuales y multimodales, radica en su capacidad para traducir información abstracta en representaciones visuales comprensibles, lo que favorece procesos de aprendizaje, análisis interdisciplinario y transferencia del conocimiento. Desde la epistemología de la mediación, estas tecnologías pueden entenderse como dispositivos hermenéuticos visuales, ya que no solo representan la realidad, sino que la reinterpretan y resignifican a través de algoritmos que priorizan ciertos rasgos, patrones y relaciones (Verbeek, 2005).

No obstante, el uso de estas plataformas en contextos científicos introduce interrogantes epistemológicas relevantes. A diferencia, de las visualizaciones tradicionales basadas directamente en datos empíricos, las imágenes generadas por IA pueden combinar elementos plausibles sin corresponder necesariamente a una representación fiel del fenómeno estudiado. Esto plantea desafíos en relación con los criterios de validación científica, la transparencia metodológica y la interpretación de los resultados visuales. Como advierte Floridi (2014), la potencia representacional de la IA puede generar una *ilusión de objetividad*, si no se explicitan los procesos y límites de generación de dichas imágenes.

Desde esta visión, la fidelidad científica de las representaciones generadas por plataformas visuales y multimodales depende en gran medida del uso epistemológicamente informado que haga el investigador. Estas herramientas, no deben ser consideradas fuentes de evidencia empírica por sí mismas, sino recursos complementarios de apoyo cognitivo y comunicativo, cuya validez reside en su coherencia con los marcos teóricos y los datos que sustentan la investigación. Tal como subraya la UNESCO (2023), el uso responsable de la IA en la ciencia exige claridad sobre el estatus epistemológico de los productos generados y una delimitación precisa de su función dentro del proceso investigativo.

De manera congruente, las plataformas visuales y multimodales constituyen un aporte significativo a la investigación académica y científica, en tanto amplían las capacidades de representación, comprensión y divulgación del conocimiento. Sin embargo, su integración debe realizarse desde una postura crítica y reflexiva, que reconozca tanto sus bondades como sus limitaciones, evitando confundir la verosimilitud visual con la validez científica. Esta comprensión resulta clave para incorporar estas tecnologías de manera ética, rigurosa y epistemológicamente responsable en la práctica investigativa contemporánea.

Entornos emergentes de productividad científica

Los entornos emergentes de productividad científica, particularmente aquellos basados en enfoques *no-code* y *low-code*, representan una evolución significativa en la infraestructura tecnológica que sostiene la investigación académica contemporánea. Plataformas como *Lovable* integran capacidades de iIAG en flujos de trabajo orientados a la automatización, articulación y optimización de tareas cognitivas y organizativas, permitiendo a los investigadores diseñar procesos, prototipos analíticos y sistemas de gestión sin requerir conocimientos avanzados de programación.

Desde una perspectiva teórica, estos entornos pueden comprenderse como parte de un proceso más amplio de digitalización cognitiva del trabajo intelectual, en el que la tecnología no solo apoya la ejecución de tareas, sino que estructura las condiciones bajo las cuales se produce el conocimiento. De acuerdo con Davenport y Kirby (2016), la automatización inteligente transforma el trabajo del conocimiento al redistribuir funciones entre humanos y sistemas algorítmicos, generando nuevas formas de colaboración y dependencia cognitiva.

En el ámbito de la investigación académica, los entornos *no-code* y *low-code* ofrecen funcionalidades clave como la organización automatizada de datos, la gestión inteligente de referencias bibliográficas, la integración de fuentes heterogéneas y la construcción rápida de prototipos analíticos o visuales. Estas capacidades reducen las barreras técnicas de acceso a herramientas avanzadas, favoreciendo una democratización de la investigación y ampliando las posibilidades de participación en procesos científicos complejos (Seely Brown & Duguid, 2002).

Uno de los principales aportes de estos entornos radica en su capacidad para externalizar y formalizar procesos cognitivos, transformándolos en flujos de trabajo configurables. Desde la teoría sociomaterial de la tecnología, esto implica que las prácticas investigativas no pueden entenderse al margen de las plataformas que las sostienen, ya que dichas plataformas configuran rutinas, prioridades y formas de validación del conocimiento (Leonardi, 2012). Así, la IA integrada en entornos de productividad científica actúa como una infraestructura cognitiva, en tanto soporta, organiza y orienta el trabajo intelectual del investigador.

Asimismo, estos entornos favorecen la experimentación metodológica, al permitir la creación rápida de modelos, simulaciones o sistemas de análisis sin una inversión significativa de tiempo o recursos técnicos. Esta flexibilidad resulta especialmente valiosa en investiga-

ciones exploratorias o interdisciplinarias, donde los marcos metodológicos se encuentran en constante ajuste. Como señalan Nonaka y Takeuchi (1999), la creación de conocimiento se ve potenciada en contextos que facilitan la conversión dinámica entre conocimiento tácito y explícito, proceso que estas plataformas tienden a acelerar.

Entre las bondades de los entornos emergentes de productividad científica destaca también su contribución a la eficiencia organizacional de la investigación, al integrar en un solo ecosistema tareas que tradicionalmente se realizaban de forma fragmentada. La automatización de procesos administrativos y técnicos libera tiempo cognitivo para actividades de mayor valor epistemológico, como la reflexión teórica, el análisis crítico y la toma de decisiones científicas fundamentadas (Zuboff, 2020).

No obstante, desde una perspectiva crítica, la consolidación de estos entornos refuerza la idea de la IA como infraestructura invisible, pero determinante de la investigación contemporánea. Al estandarizar flujos de trabajo y criterios de organización del conocimiento, estas plataformas pueden influir en las formas legítimas de investigar, planteando desafíos en términos de dependencia tecnológica, transparencia metodológica y autonomía epistémica. Reconocer esta dimensión infraestructural resulta clave para una integración reflexiva y responsable de estos entornos en la práctica científica.

Cabe decir que los entornos emergentes de productividad científica basados en IAG no deben entenderse únicamente como herramientas de apoyo técnico, sino como marcos socio-técnicos que reconfiguran las prácticas, ritmos y estructuras de la investigación académica. Su análisis permite comprender cómo la IA se consolida como una infraestructura cognitiva central en la producción de conocimiento, con implicaciones profundas para la organización, la ética y la epistemología de la ciencia contemporánea.

Discusión: implicaciones epistemológicas y metodológicas

La incorporación de la IAG en la investigación académica y científica supone una transformación profunda en las condiciones epistemológicas de producción del conocimiento. Uno de los cambios más significativos, es la redefinición del rol del investigador, quien deja de ser concebido exclusivamente como el autor individual del discurso científico para asumir una función más compleja como curador epistémico. Este nuevo rol, implica evaluar críticamente los aportes generados por sistemas de IA, contextualizarlos dentro de marcos teóricos específicos y validar su coherencia metodológica y científica.

Desde una perspectiva epistemológica, esta transformación cuestiona concepciones tradicionales de la autoría y la originalidad científica. Como señalan Polanyi y Sen (2009), todo conocimiento humano incorpora una dimensión tácita que no puede ser completamente formalizada ni automatizada. En este sentido, aunque la IAG puede producir textos, esquemas o representaciones plausibles, la responsabilidad epistémica del investigador reside en interpretar, seleccionar y dotar de sentido dichos productos, integrándolos en un proceso reflexivo que no puede ser delegado a sistemas algorítmicos.

Esta reconfiguración, del rol investigador exige el desarrollo de nuevas competencias epistemológicas y metodológicas. Entre ellas destaca la alfabetización en IA, entendida no solo como el conocimiento funcional de las herramientas, sino como la capacidad de comprender sus supuestos, límites y efectos sobre la construcción del conocimiento. De acuerdo con la Organización y Cooperación para el Desarrollo Económico (OCDE, 2021), la alfabetización en IA es una competencia clave para los profesionales del conocimiento, ya que permite un uso crítico y responsable de tecnologías que influyen directamente en la toma de decisiones y en la producción intelectual.

Asimismo, la formulación crítica de instrucciones (*prompts*) emerge como una práctica metodológica relevante. Esta habilidad, implica traducir problemas teóricos complejos en instrucciones claras, reflexivas y epistemológicamente fundamentadas, lo cual requiere un dominio conceptual previo por parte del investigador. En este sentido, la interacción con la IA no sustituye el pensamiento científico, sino que lo pone a prueba, evidenciando la necesidad de claridad conceptual y rigor argumentativo (Schön & DeSanctis 1986).

No obstante, la integración de la IAG en la investigación académica también conlleva riesgos epistemológicos y metodológicos significativos. Entre los más relevantes se encuentra la dependencia cognitiva, entendida como la delegación excesiva de procesos de análisis, síntesis y argumentación en sistemas automatizados. Por su parte, Salomón (1992) advierte que cuando las tecnologías cognitivas asumen funciones centrales del pensamiento, existe riesgo de un debilitamiento de las habilidades reflexivas si no se mantiene una interacción consciente y crítica.

Otro riesgo identificado es la homogeneización del discurso científico, derivada del uso generalizado de modelos entrenados con *corpus* dominantes que tienden a reproducir estilos argumentativos, enfoques teóricos y paradigmas hegemónicos. Esta tendencia, puede limitar la diversidad epistémica y metodológica, afectando particularmente a enfoques críti-

cos, perspectivas periféricas o saberes emergentes. Desde la sociología del conocimiento científico, Becher y Trowler (2001) señalan que las comunidades académicas se caracterizan por culturas disciplinares diversas, las cuales pueden verse erosionadas por procesos de estandarización discursiva.

Frente a estos desafíos, se propone un modelo de integración epistemológicamente responsable de la IAG, estructurado en tres dimensiones interrelacionadas:

1. Dimensión cognitiva, orientada a preservar y fortalecer la autonomía intelectual del investigador, promoviendo el uso de la IA como apoyo reflexivo y no como sustituto del pensamiento científico.

2. Dimensión metodológica, centrada en la transparencia del uso de la IA, la explicitación de sus funciones dentro del proceso investigativo y la coherencia entre los productos generados y los marcos teóricos adoptados.

3. Dimensión ética, enfocada en la responsabilidad, la justicia epistémica y la evaluación crítica de los sesgos, impactos y límites de la IA en la producción del conocimiento.

Este modelo reconoce que la IAG no es un elemento externo o accesorio, sino un actor estructural en la ecología contemporánea de la investigación científica. Su integración responsable depende, en última instancia, de la capacidad del investigador para ejercer una reflexividad epistemológica constante, orientada a salvaguardar el rigor, la diversidad y la integridad del conocimiento académico.

Línea de investigación propuesta: epistemología aplicada de la inteligencia artificial en la investigación científica

A partir del análisis teórico y comparativo desarrollado, se propone la línea de investigación denominada *Epistemología aplicada de la inteligencia artificial en la investigación científica*, orientada a examinar de manera crítica, sistemática y multidimensional el impacto de la inteligencia artificial en los procesos contemporáneos de producción, validación y comunicación del conocimiento académico y científico.

Esta línea se inscribe en el cruce entre la epistemología contemporánea, los estudios sociales de la ciencia y la tecnología y la metodología de la investigación, reconociendo que la IA no opera únicamente como una herramienta técnica, sino como un mediador epistémico que reconfigura las formas de conocer, argumentar y legitimar saberes. En este

sentido, la IA se comprende como parte de una infraestructura cognitiva que incide tanto en los procesos individuales de pensamiento como en las dinámicas colectivas de las comunidades científicas.

Desde un proceso epistemológico, esta línea de investigación problematiza los supuestos tradicionales sobre la neutralidad de los instrumentos científicos. Diversos enfoques contemporáneos han señalado que toda tecnología de conocimiento incorpora valores, supuestos y orientaciones implícitas que influyen en los resultados cognitivos (Bunge, 2012; Latour, 2008). La IAG, al intervenir en tareas como la formulación de problemas, la síntesis teórica o la redacción académica, introduce nuevas formas de mediación que requieren ser analizadas con rigor conceptual y metodológico.

Conclusiones

La IAG constituye uno de los cambios más significativos en la historia reciente de la investigación académica y científica. Los resultados de este estudio permiten afirmar que su impacto trasciende ampliamente la dimensión instrumental o técnica, para situarse en el núcleo mismo de los procesos epistemológicos que configuran la producción, validación y comunicación del conocimiento. En este sentido, la IAG no solo optimiza tareas investigativas, sino que redefine las condiciones bajo las cuales se construye el saber científico en la contemporaneidad.

Uno de los principales hallazgos de la investigación es la identificación de la IAG como mediador epistémico activo, y no como un recurso neutral. El análisis teórico-documental, mostró que plataformas cómo los modelos de lenguaje de gran escala y los sistemas multimodales influyen de manera directa en la formulación de problemas, la organización conceptual, la jerarquización de argumentos y la selección de marcos teóricos. Esta mediación, transforma las prácticas cognitivas del investigador y exige una relectura crítica de nociones clásicas como autoría, originalidad y autonomía intelectual.

Asimismo, queda en evidencia que la incorporación de la IAG modifica los roles tradicionales del investigador, quien deja de ser el único productor del discurso científico para asumir una función de curaduría epistémica. Este nuevo rol implica evaluar la pertinencia, coherencia y validez de los contenidos generados algorítmicamente, así como contextualizarlos dentro de tradiciones teóricas y metodológicas específicas. En consecuencia, la competencia investigativa ya no se define únicamente por el dominio disciplinar, sino también por la capacidad de interactuar críticamente con sistemas de IA.

Otro hallazgo relevante se relaciona con la dimensión colectiva del conocimiento mediado por IA. Desde la epistemología social, se constató que los modelos generativos incorporan patrones discursivos, consensos científicos y sesgos provenientes de comunidades epistémicas amplias. Esto refuerza, la idea de que la IA participa en la producción social del conocimiento, pero también pone en evidencia riesgos asociados a la homogeneización del discurso científico, la reproducción de sesgos estructurales y la posible reducción de la diversidad epistemológica.

A partir de estos hallazgos, el estudio propone una integración epistemológicamente responsable de la IAG, estructurada en tres dimensiones interdependientes: cognitiva, metodológica y ética. Esta propuesta reconoce el potencial transformador de la IA, pero subraya la necesidad de establecer criterios claros para su uso crítico, evitando tanto la dependencia cognitiva como la adopción acrítica de sus resultados.

Finalmente, este trabajo concluye que la IAG no representa una amenaza inevitable para la investigación académica, sino un desafío epistemológico de alta complejidad que requiere marcos teóricos renovados, formación investigativa específica y principios éticos sólidos. La línea de investigación propuesta —*Epistemología aplicada de la inteligencia artificial en la investigación científica*— se presenta, así, como una vía necesaria para comprender, orientar y regular el papel de la IA en la ciencia contemporánea, contribuyendo a una producción de conocimiento más reflexiva, responsable y socialmente pertinente.

Referencias

Becher, T., y Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and fthe cultures of disciplines* (2nd ed.). Open University Press.

Boden, M. A. (2017). *Inteligencia artificial*. Turner.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). On the opportunities and risks of foundation models [Informe técnico]. *arXiv*. https://arxiv.org/abs/2108.07258

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901. https://arxiv.org/abs/2005.14165

Bunge, M. (2012). *Epistemología y metodología I: Explorando el mundo*. Springer Science & Business Media.

Cairo, A. (2016). *El arte veraz: Datos, gráficos y mapas para la comunicación*. New Riders Publishing.

Davenport, T. H., & Kirby, J. (2016). *Solo los humanos deben postularse: Ganadores y perdedores en la era de las máquinas inteligentes* (Vol. 1). Harper Business.

Feenberg, A. (2002). *Transforming Technology: A Critical Theory Revisited*. Oxford University Press.

Floridi, L. (2014). *La cuarta revolución: cómo la infosfera está transformando la realidad humana*. Oxford University Press.

Floridi, L. (2024). *Ética de la inteligencia artificial.* Herder.

García-Peñalvo, F. J. (2024). Inteligencia artificial generativa y educación: Un análisis desde múltiples perspectivas. *Education in the Knowledge Society* (*EKS*), 25, e31942. https://doi.org/10.14201/eks.31942

García Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, *27*(1), 9–39. https://doi.org/10.5944/ried.27.1.37716

Hayles, N. (2017). *Unthought: The power of the cognitive nonconscious*. University of Chicago Press. https://doi.org/10.7208/9780226447919

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences.* SAGE. Publications Ltd.

Knorr Cetina, K. (2005). *La fabricación del conocimiento: Un ensayo sobre el carácter constructivista y contextual de la ciencia*. Universidad Nacional de Quilmes.

Latour, B. (2008). *Re-ensamblar lo social. Una introducción a la teoría del actor-red*. Manantial.

Leonardi, P. M. (2012). Materiality, sociomateriality, and socio-technical systems: What do these terms mean? How are they different? Do we need them. *Materiality and organizing: Social interaction in a technological world*, *25*(10), 1093.https://doi.org/10.1093/acprof:oso/9780199664054.003.0002

Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University.

Nonaka, I., & Takeuchi, H. (1999). *La organización creadora de conocimiento: cómo las compañías japonesas crean la dinámica de la innovación* (Vol. 1). Oxford University Press.

OCDE. (2021). *Artificial intelligence, data and digital transformation in education*. OECD Publishing.

Polanyi, M., & Sen, A. (2009). *The tacit dimension.* University of Chicago Press.

Russell, S., & Norvig, P. (2004). *Inteligencia Artificial: un enfoque moderno* (2.ª ed.). Pearson Educación.

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. Global Edition.

Salomon, G. (1992). *Las tecnologías de la información y la educación: Nuevas oportunidades, nuevos problemas*. Paidós.

Schön, D., & DeSantics, V. (1986). El profesional reflexivo: Cómo piensan los profesionales en acción. *The Journal of Continuing Higher Education*, *34*(3). https://doi.org/10.1080/07377366.1986.10401080

Seely Brown, J., & Duguid, P. (2002). *La vida social de la información*. Pearson Educación.

Simondon, G. (2008). *El modo de existencia de los objetos técnicos*. Prometeo.

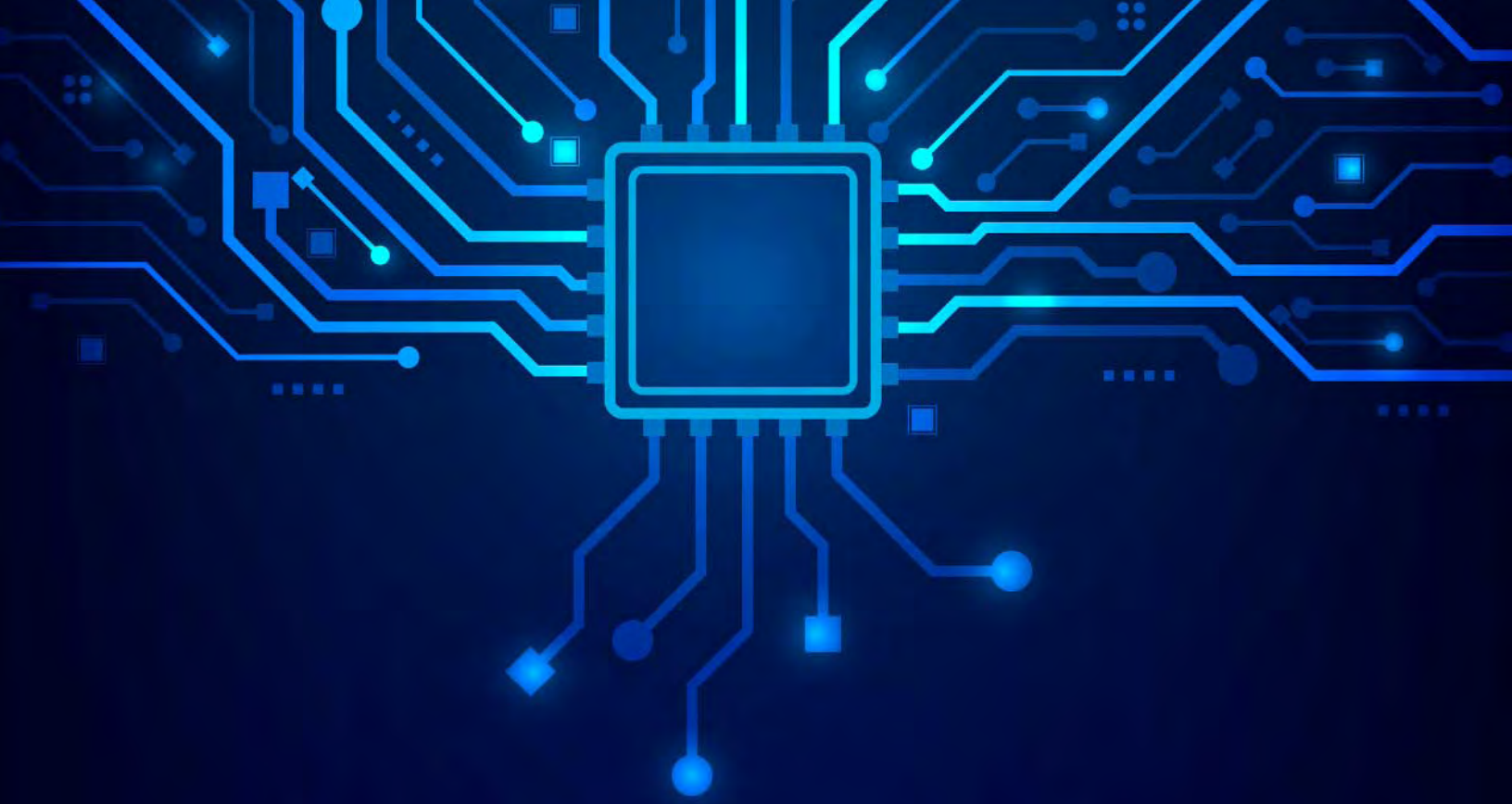
Tufte, E. (2006). *La visualización de la información cuantitativa*. Graphics Press.

UNESCO. (2023). *Marco ético para la inteligencia artificial en la educación*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://unesdoc.unesco.org/ark:/48223/pf0000381564

Verbeek, P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press. https://philpapers.org/rec/VERWTD

Verbeek, P. (2011). *Moralizar la tecnología: Comprender y diseñar la moralidad de las cosas*. University of Chicago Press. https://philpapers.org/rec/VERMTU-2

Zuboff, S. (2020). *La era del capitalismo de vigilancia. La lucha por un futuro humano frente a las nuevas fronteras de poder*. Paidós.



25.

Análisis documental de la adopción y humanización de asistentes digitales basados en inteligencia artificial

Karina Guadalupe Cortina Calderón
Universidad Autónoma de Tamaulipas, México
ORCID: /0009-0008-8153-3306

Nallely Guadalupe Hernández Hernández
Universidad Autónoma de Tamaulipas, México
ORCID: 0000-0003-1249-0118

Mónica Lorena Sánchez Limón
Universidad Autónoma de Tamaulipas, México
ORCID: 0000-0002-0671-0076

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.25

25.

Análisis documental de la adopción y humanización de asistentes digitales basados en inteligencia artificial

Introducción

Desde sus primeras apariciones, la inteligencia artificial (IA) ha evolucionado pasando a ser de promesa computacional a un agente social que moldea experiencias, percepciones y decisiones de millones de usuarios (Cortina et al., 2024). Aunque en sus inicios no se utilizaba este término, las investigaciones relacionadas con esta tecnología se remontan a la década de 1950, coincidiendo con la aparición de las primeras computadoras, por lo que, a partir de entonces, la IA ha fortalecido su desarrollo y su influencia en la vida diaria se ha hecho cada vez más evidente (Makridakis, 2017).

Esto ha propiciado que la industria de la IA alcance una valoración cercana a los cien mil millones de dólares. Además, sus proyecciones son altamente favorables. De acuerdo con estimaciones de *Statista*, se anticipa un crecimiento exponencial que podría llevar a este mercado a aproximarse a los 2 billones de dólares para el año 2030. Este desarrollo implicaría su expansión en múltiples sectores e integración en procesos como las cadenas de suministro, el *marketing*, la manufactura, la investigación y diversas estructuras empresariales (Thormundsson, 2023).

A partir de la amplia diversidad de dispositivos en los que se integra la IA, emerge el concepto *asistente digital* (AD), entendido como una tecnología avanzada sustentada en IA. En este sentido, Brill et al. (2019) y Marikyan et al. (2022) lo describen como sistemas complejos capaces de operar en distintos entornos. Este concepto engloba una variedad de dispositivos interactivos que responden a comandos de voz o texto, como los asistentes virtuales por voz y los *chatbots*, entre otros (Pantano & Pizzi, 2020; Ramadan, 2021).

Debido a su rápida difusión en la última década, los asistentes digitales, por ejemplo, de voz o agentes conversacionales, ha facilitado a los usuarios expresar sus requerimientos

de manera más inmediata y natural (Marriott & Pitardi, 2021). Esta tendencia ha sido aprovechada en distintas estrategias de *marketing*, en las que marcas como *Mr. Clean*, *Tony the Tiger* y *Energizer Bunny* ejemplifican productos dotados de atributos humanos, creados con el propósito de facilitar la interacción con los consumidores, mejorar sus percepciones e influir en sus decisiones de compra, lo que contribuye al incremento en las ventas (Huang et al., 2024; Malhotra & Ramalingam, 2023).

Esto ha dado lugar a investigaciones que demuestran que la antropomorfización de los productos (Aggarwal & McGill, 2007), es decir, atribuirles características humanas, influye positivamente en la percepción de los consumidores. En particular, el uso de pronombres personales o la referencia en primera persona, en lugar de la tercera, al describir un producto tiende a generar evaluaciones más favorables por parte del público.

No obstante, resulta fundamental comprender y aplicar el antropomorfismo de manera equilibrada, ya que una sobrestimación o subestimación de sus efectos puede generar percepciones distorsionadas en los usuarios. Por ello, al considerar las preferencias y capacidades cognitivas de las personas, se vuelve esencial analizar cómo la incorporación de rasgos humanos en avatares y agentes virtuales influye en su percepción y experiencia. En esta línea, Sharma et al. (2022) señalan que una comprensión profunda de estos factores permite diseñar interacciones más eficaces y satisfactorias entre humanos y computadoras, favoreciendo una mayor aceptación de la tecnología.

Por ello, el objetivo del presente estudio es comprender las líneas de investigación que se encuentran en el marco de los asistentes digitales, a fin de lograr responder las preguntas de investigación ¿Cuáles y como son las líneas de investigación en el escenario de asistentes digitales?

Revisión de la literatura

La investigación en el campo de los AD ha experimentado un crecimiento significativo, en diversas disciplinas (Ramadán, 2021). Ello, en función de que asistentes digitales han adquirido una notable relevancia a nivel global, debido a su amplia disponibilidad en diversas modalidades que abarcan desde dispositivos móviles hasta sistemas de hogar inteligente, así como su integración en servicios de atención al cliente y en aplicaciones de búsqueda y recomendación personalizada, entre otros (IMMNUNE, 2023; Knight, 2012).

Por ello, los AD se han convertido en una pieza clave debido a su capacidad para simplificar tareas cotidianas, mejorar la eficiencia y la experiencia del usuario, impulsando así el continuo avance de la tecnología en diferentes áreas de la vida moderna (Sáiz et al., 2020). Esta importancia radica en su habilidad para entender el lenguaje natural, procesar datos de manera inteligente y adaptarse a las necesidades de los usuarios a través de acciones como movilidad, interacción de voz, delegación de tareas y análisis rápido de datos (Griol & Callejas, 2016). Estas capacidades han llevado a su rápida adopción y a un mayor desarrollo en el campo de la inteligencia artificial.

A pesar de este avance, persisten ciertas brechas en la literatura que requieren ser identificadas y abordadas mediante nuevos estudios. Dentro de las ciencias sociales, las principales líneas de investigación que versan sobre los AD intentan comprender las motivaciones que llevan al individuo a interactuar con dichas tecnologías (Ferreira, 2021). Por lo que, las principales líneas de investigación que estudian este fenómeno se pueden agrupar en dos corrientes principales (Ferreira, 2021; Marriott & Pitardi, 2021; McLean et al., 2021).

La primera, se ha enfocado en analizar la intención de comportamiento de los consumidores hacia el uso de asistentes digitales aplicando teorías relacionadas con la adopción de tecnologías (Moriuchi, 2021). Esta corriente de investigación ha sido ampliamente utilizada en literatura de comercio electrónico, comercio móvil y redes sociales, y también se han adoptado en investigaciones sobre IA y AD. Estos enfoques teóricos respaldan la influencia que ejercen factores como la utilidad percibida y la facilidad de uso en las intenciones de adopción tecnológica por parte de los individuos (Marriott & Pitardi, 2021).

La segunda corriente de la literatura ha dado lugar a la investigación de la adopción de AD a través de las relaciones construidas entre estos y los consumidores, misma que se da por los atributos humanos que perciben en ellos (Marriott y Pitardi, 2021; McLean et al., 2021). Esta línea considera los aspectos sociales y la capacidad de los asistentes digitales para ser interactivo y socialmente agradable, generando una sensación de conexión interpersonal que influye directamente en la satisfacción de los usuarios y en su disposición a seguir utilizando estas tecnologías, especialmente cuando las interacciones se vuelven más frecuentes y habituales en el entorno doméstico o personal (Han & Yang, 2018).

Dado el contexto que se vive en donde los avances tecnológicos se encuentran en un dinamismo continuos, se requieren más estudios e investigaciones que permitan comprender de qué manera los AD basados en IA impactan y moldean el comportamiento de los

consumidores, así como su contribución en la transformación de la interacción y la dinámica entre los consumidores y las empresas (Van Doorn et al., 2016). Generando que la literatura respecto a estas dos corrientes haya sido criticada, proponiendo la necesidad de combinar ambas líneas de investigación para obtener una investigación más clara y concreta respecto al uso de los AD (Marriott & Pitardi, 2021).

Diseño metodológico

La presente investigación se enmarca en un enfoque documental, sustentado en la recopilación, revisión y análisis de información proveniente de diversas fuentes académicas y científicas, con el fin de fundamentar y estructurar la metodología del problema de estudio. A partir de este proceso, se adopta un alcance descriptivo orientado a identificar, organizar y caracterizar las principales líneas de investigación vinculadas con el tema, lo que permite una comprensión integral de sus enfoques, tendencias y aportaciones dentro del campo de conocimiento. Para su desarrollo, se seleccionaron y analizaron los dieciocho artículos más relevantes en torno al estudio de los asistentes digitales.

Resultados

Las Tablas 1 y 2, exponen los resultados del análisis documental. A partir de la revisión de la literatura se encontró que existen dos líneas de investigación con amplios resultados, mismas que en los recientes años han versado sobre los estudios de asistentes digitales impulsados por inteligencia artificial, estos son denominados para efectos del cumplimiento del objetivo de la presente investigación características tecnológicas del objeto y atributos humanos incorporados, mismos que se explicarán a continuación.

Como se mencionó anteriormente, la primera línea de investigación detectada corresponde a las A) *características tecnológicas del objeto*, partiendo desde el estudio de Chuah et al., (2021) quienes proponen que una de las clasificaciones de las características de los AD son aquellas funciones con semejanza a la tecnología que llevan a los consumidores a aceptar y utilizar un nuevo producto tecnológico. Esta corriente se centran en la aplicación de marcos teóricos establecidos, como el Modelo de Aceptación Tecnológica (TAM por sus siglas en inglés) (Davis et al., 1989) y la Teoría Unificada de Aceptación y Uso de la Tecnología (UTAUT por sus siglas en inglés) (Venkatesh et al., 2003), que han sido validados previamente en contextos de adopción y utilización de conceptos tecnológicos convencionales.

Tabla 1

Revisión de la literatura de características tecnológicas del objeto

Autor	Objetivo	Tipo de asistente	País	Variables	Teoría
Wang et al. (2023)	Examinar cómo se puede hacer para que la IA sea más efectiva y rentable en el comercio electrónico y cómo los empresarios pueden hacer que la tecnología de IA les ayude a lograr sus objetivos comerciales.	IA al sitio web de una empresa de comercio electrónico (*chatbot*, asistentes de voz)	Pakistán	Actitud de uso Intención conductual Uso real Facilidad de uso percibida Utilidad percibida Normas subjetivas Confianza	Modelo de Aceptación de Tecnología (TAM)
Basak et al. (2015)	Identificar los factores que afectan la intención de utilizar la tecnología de asistente digital personal (PDA)	Asistente digital personal (PDA)	Turquía	Norma subjetiva Facilidad de uso percibida Innovación personal Utilidad percibida Autoeficacia informática Intención de uso Disfrute percibido	Modelo de Aceptación de Tecnología (TAM) extendido
Zou & Huang (2023)	Investigar la aceptación escrita de los estudiantes de doctorado hacia ChatGPT y los factores que influyen en ella.	Chatbot (ChatGPT)	China	Facilidad de uso percibida Utilidad de uso percibida Actitud de uso Intención conductual de uso.	Modelo de Aceptación de Tecnología (TAM)

Tabla 1

Revisión de la literatura de características tecnológicas del objeto

Autor	Objetivo	Tipo de asistente	País	Variables	Teoría
Jain et al., (2022)	Explorar los factores que influyen en el uso de herramientas habilitadas por IA	Herramientas basadas en IA	India	Condiciones facilitadoras Influencia social Uso Aversión a la IA Expectativa de desempeño Expectativa de esfuerzo Colaboración	Teoría unificada de aceptación y uso de la tecnología (UTAUT)
Roy et al. (2022)	Medir la intención de adoptar robots basados en IA en el sistema educativo de las universidades indias	Robots basados en IA	India	Facilidad de uso percibida Utilidad percibida Actitud Confianza Intención de adoptar Malestar Inseguridad	Modelo TAM, Teoría del Comportamiento Planificado (TPB) y el índice de preparación tecnológica (TRI)

Tabla 1

Revisión de la literatura de características tecnológicas del objeto

Autor	Objetivo	Tipo de asistente	País	Variables	Teoría
Cao et al. (2021)	Considerar los factores positivos y negativos que influyen colectivamente en las actitudes y las intenciones de comportamiento hacia el uso de la IA.	IA para la toma de decisiones	Reino Unido	Influencia de los pares Condiciones facilitadoras Intención de utilizar Expectativas de rendimiento Expectativa de esfuerzo Actitud hacia el uso Preocupación por el bienestar personal Preocupación por el desarrollo personal Gravedad percibida Susceptibilidad percibida Amenaza percibida Actitud	Teoría unificada de aceptación y uso de la tecnología (UTAUT), Teoría de evitación de amenazas tecnológicas (TTAT), Modelo integrado de aceptación y evitación de IA (IAAAM)

Tabla 1

Revisión de la literatura de características tecnológicas del objeto

Autor	Objetivo	Tipo de asistente	País	Variables	Teoría
Balakrishnan & Dwivedi (2021)	Investigar los factores que generan resistencia y actitud hacia los asistentes de voz de IA	Experimento comprobado con modelos de ecuaciones estructurales	India	Costo hundido Evitar el arrepentimiento Inercia Valor percibido Adopción Costos de cambio Amenaza percibida Facilidad de uso Utilidad percibida Actitud Resistencia a la adopción	Marco de doble factor y Modelo de Aceptación de Tecnología (TAM)
Silva et al. (2023)	Investigar los factores que influyen en la intención de los clientes de reutilizar los servicios basados en chatbot	*Chatbot*	Portugal	Confianza Satisfacción del usuario Utilidad percibida Actitud hacia el uso Intenciones de reutilización Presencia social percibida Presencia social Facilidad de uso percibida Norma subjetiva	Modelo de Aceptación de Tecnología (TAM) y TAM 2

Tabla 1
Revisión de la literatura de características tecnológicas del objeto

Autor	Objetivo	Tipo de asistente	País	Variables	Teoría
Lubbe & Ngoma (2021)	Examinar cómo la facilidad de uso percibida, la alegría y la utilidad percibidas de los chatbots influyen en la experiencia del cliente y cómo estas experiencias influyen en la satisfacción de los usuarios en un contexto de mercado emergente	*Chatbot*	Sudáfrica	Utilidad percibida Facilidad de uso percibida Alegría percibida Experiencia en tecnología de autoservicio Satisfacción con la tecnología de autoservicio	Modelo TAM extendido

Estos marcos se derivan de las corrientes de pensamiento propuestas por Martin Fishbein e Icek Ajzen, como lo son la Teoría de Acción Razonada (TRA por sus siglas en inglés) (Ajzen & Fishbein, 1972) y la Teoría del Comportamiento Planeado (TPB, por sus siglas en inglés) (Ajzen & Fishbein, 1980). Estos modelos examinan factores como la utilidad, facilidad de uso, expectativas de esfuerzo y rendimiento, influencia social, condiciones facilitadoras, confianza, entre otros, así como sus consecuencias en las actitudes y comportamientos del usuario (Medina et al., 2021; Wang et al., 2023). De aquí que nace el interés por comprender las características tecnológicas que influyen en la intención de uso por parte de los usuarios, así como los factores adicionales que pueden moldear la adopción y la experiencia de uso de estas nuevas tecnologías en constante evolución.

En la literatura de esta línea de investigación, se han explorado y analizado esas características específicas que motivan o desmotivan a los usuarios en la adopción y uso de diversas tecnologías, incluidos los asistentes digitales o aquellos objetos asistidos por inteligencia artificial. Como se mencionó anteriormente, a medida que las tecnologías evolucionan y se vuelven más sofisticadas, surgen nuevos factores que influyen en la intención de uso por parte de los usuarios (Venkatesh et al., 2012).

Las investigaciones que se exponen en la Tabla 1 están basadas en modelos teóricos anteriormente mencionados como el Modelo de Aceptación de Tecnología (TAM) y la

Teoría Unificada de Aceptación y Uso de Tecnología (UTAUT) utilizados en estos estudios para investigar la aceptación y adopción de tecnologías de inteligencia artificial, tales como chatbots, asistentes de voz y herramientas habilitadas para IA.

Dentro de estos estudios, se examinan las características que influyen en la intención de usar tecnologías de IA en diferentes contextos, como comercio electrónico, atención médica, educación superior y toma de decisiones organizacionales (Basak et al., 2015; Jain et al., 2022; Wang et al., 2023; Zou y Huang, 2023).

En estas investigaciones, se han identificado una serie de características comunes que influyen en la actitud y la intención de uso de los usuarios. Estas incluyen algunas, como la utilidad percibida, la facilidad de uso percibida, las normas subjetivas o influencia social, la confianza y el disfrute percibido (Basak et al., 2015; Lubbe y Ngoma, 2021; Silva et al., 2023). Sin embargo, algunos estudios han ido más allá al incorporar nuevos constructos, como el costo hundido, la amenaza percibida, el cinismo sobre la privacidad y la aversión a la IA, para obtener una comprensión más completa de la aceptación y el rechazo de la IA. Además, se ha destacado la importancia de características como la satisfacción del usuario, inseguridad, aversión a la IA y la presencia social percibida en la aceptación específica de chatbots o asistentes de voz, subrayando la complejidad y la multidimensionalidad de este fenómeno (Balakrishnan y Dwivedi, 2021; Cao et al., 2021; Jain et al., 2022).

Por otro lado, la línea de investigación B) *atributos humanos incorporados*, hace alusión a como las tecnologías asemejan a características humanas. En este ámbito, predominan teorías como la teoría del Valle del Valle Inquietante (Mori et al., 2012), la Interacción Humano-Computadora (Sharples, 1996), Paradigma CASA (computadoras como actores sociales) (Reeves y Nass, 1996), entre otras, lo cual sugiere que los usuarios desarrollan interacciones personales con la tecnología, sintiendo que los mensajes que reciben han sido diseñados especialmente para ellos, de manera similar a como lo haría un amigo o una pareja cercana (Cortina et al., 2024).

Tabla 2

Revisión de la literatura de los atributos humanos incorporados

Autor	Objetivo	Tipo de asistente	País	Variables	Teoría
(Belanche et al., 2020)	Complementar la investigación sobre la semejanza humana de los robots con investigaciones de los factores que los gerentes de servicios deben elegir para los robots de servicios implementados en su entorno de servicios	Robots de servicios	NM	Estética (antropomorfismo, género, etnicidad y cultura) Notificación de robot Manipulabilidad Proactividad Afectar Formalidad Características del cliente Preparación tecnológica Edad Género Cultura Rasgos de personalidad Nivel de cliente	Semejanza humana de los robots basada en tres categorías: diseño del robot, las características del cliente y las características del encuentro con el servicio
(Lau et al., 2018)	Estudiar las razones de las personas a favor y en contra de la adopción de parlantes inteligentes, sus percepciones y preocupaciones sobre la privacidad, y sus comportamientos de búsqueda de privacidad	Asistentes de voz (*Amazon Echo* y *Google Home*)	NM	Confianza Privacidad	Interacción Humano-Computadora
(De Cicco et al., 2020)	Investigar más a fondo cómo la perciben los usuarios y cuáles son las variables que potencian una actitud positiva hacia esta tecnología	*Chatbots* orientados a las tareas	Italia	Presencia social Interacción Disfrute percibido Confianza Actitud	interacción persona-computadora (HCI), presencia social

(Song & Shin, 2024)	Investigar el efecto de la humanización del *chatbot* en la percepción de inquietud, confianza y la intención de comportamiento de los usuarios	*Chatbots* de comercio electronico	Países bajos	Interacción con *chatbots* hiperrealistas y animados Sensación de inquietud que la interacción con *chatbots* de dibujos animados y quietos Familiaridad Apariencia humana del *chatbot* Confianza Intención de compra Voluntad de reutilizar el *chatbot* en el futuro	Teoría del valle inquietante (UVE) y paradigma CASA (Las computadoras son actores sociales), Teoría de la Acción Razonada y Teoría del Comportamiento Planificado
(Ciechanowski et al., 2019)	Examinar a fondo el carácter del proceso de interacción entre humanos y no humanos.	*Chatbots* texto y *chatbots* avatar	NM	Efectos extraños Efectos negativos *Chatbot* con avatar *Chatbot* de texto Datos psicofisiológicos	Interacción Humano-Computadora, Efecto Valle Inquietante, Teoría Comportamiento Planeado y Presencia Social
(Li et al., 2021)	Comprender cómo las características sociales pueden beneficiar las interacciones entre humanos y *chatbots* e identificar los desafíos y estrategias para diseñarlas.	*Chatbots* de servicios de turismo	China	Calidad del *chatbot* Confirmación posterior al uso Calidad del servicio Ansiedad tecnológica Confirmación posterior al uso del servicio Satisfacción Intención de continuidad del uso	Teoría del valle inquietante (UVE), Humanidad Tecnológica (presencia social) y Personificación

(Balakrishnan y Dwivedi, 2021)	Investigar el papel de la actitud tecnológica y los atributos de la IA en la mejora de la intención de compra a través de asistentes digitales.	Asistentes digitales	India	Facilidad de uso percibida Utilidad percibida Actitud antropomorfismo percibido Intención de compra Inteligencia percibida Animacidad percibida Experiencia de uso	Modelo TAM, Teoría del Valle Inquietante, Teoría de la Mente, Psicología del Desarrollo y Teoría de la Psicología Cognitiva.
(Bartneck et al., 2009)	Enfatizar la necesidad de herramientas de medición estandarizadas para la interacción entre humanos y robots	Robots de servicios	Japón	Antropomorfismo Animacidad Simpatía Inteligencia percibida Seguridad percibida	Interacción Humano-Robots
(McLean et al., 2021)	Descubrir los impulsores clave del compromiso de la marca del consumidor a través de los asistentes de voz	Asistente de Voz Doméstico (*Echo* y *Alexa*)	NM	Presencia social Engagement con la marca Atracción social Engagement con la marca Utilidad de uso percibida Facilidad de uso percibida Beneficios utilitarios Búsqueda de interacciones hedonistas Confianza Intención de uso Intención de uso Intención de compra	Engagement con la marca, Presencia Social y Modelo TAM

Estos estudios exploran cómo las características de estas tecnologías, que imitan atributos humanos como el antropomorfismo, tono de voz, personalidad, sociabilidad, inteligencia, movimiento (animación), rasgos faciales o habilidades mentales, estética, familiaridad, entre otros que impactan en la percepción y la aceptación de los usuarios hacia estas innovaciones tecnológicas (Moussawi &y Koufaris, 2019; Song & Shin, 2024).

A pesar de que esta línea de investigación se ha centrado en explorar las respuestas emocionales (afecto) asociadas con este efecto, sigue existiendo un conocimiento limitado sobre elementos como la percepción real de las personas hacia los objetos humanoides, cuáles son los factores específicos que contribuyen a este denominado valle inquietante, así como las razones que detonaron la aparición del su efecto, mismo que no ha sido explicado en su totalidad (Ciechanowski et al., 2019).

Conclusiones

La presente investigación permitió reconocer que el estudio de los AD se ha consolidado como un campo relevante y en expansión dentro de diversas disciplinas, particularmente en las ciencias sociales y administrativas. A partir del análisis documental, se observa que la adopción de estas tecnologías responde a la interacción de factores técnicos, cognitivos, sociales y emocionales, los cuales influyen en la forma en que los usuarios se relacionan con sistemas basados en IA (Ferreira, 2021).

En este sentido, se identificaron dos grandes líneas de investigación que estructuran el estudio de los asistentes digitales. La primera se enfoca en las características tecnológicas, sustentándose en modelos como el TAM y el UTAUT, los cuales explican la adopción a partir de variables como la utilidad percibida, la facilidad de uso, la confianza y la influencia social (Venkatesh et al., 2003). Estos enfoques han mostrado solidez para explicar la intención de uso y han incorporado nuevos constructos que amplían la comprensión del fenómeno (Wang et al., 2023; Medina et al., 2021).

Por otra parte, la segunda línea de investigación resalta la relevancia de los atributos humanos en los asistentes digitales, especialmente aquellos vinculados con el antropomorfismo y la interacción social percibida. Estos elementos favorecen la cercanía entre el usuario y la tecnología, influyendo en la satisfacción, la confianza y la continuidad de uso (Han & Yang, 2018).). Asimismo, la aparición de teorías más recientes, demuestran las nuevas percepciones que los individuos están teniendo con la incorporación de estos atributos humanos

en las actividades tradicionales de comercio, atención al cliente, entre otras. Teorías como el Paradigma CASA (Reeves & Nass, 1996) muestra que los usuarios interactúan con las tecnologías mediáticas de manera intrínsecamente social y espontánea. De igual manera, el efecto del *Valle Inquietante* representa un aspecto clave en la percepción de los usuarios, ya que incide en sus respuestas emocionales frente a tecnologías con rasgos humanoides (Mori et al., 2012; Ciechanowski et al., 2019).

De igual forma, los resultados evidencian que la adopción y uso de AD constituye un fenómeno complejo y multidimensional, en el que convergen factores que impulsan o limitan su aceptación. Variables como la satisfacción, el disfrute y la confianza favorecen su uso, mientras que elementos como la inseguridad, la aversión a la IA y la percepción de amenaza inciden en las actitudes y comportamientos de los usuarios (Cao et al., 2021; Jain et al., 2022; Silva et al., 2023).

Finalmente, se identifica una oportunidad relevante para futuras investigaciones orientadas a integrar ambas líneas de estudio, articulando los enfoques tecnológicos con los aspectos humanos y sociales. Esta integración permitirá generar una comprensión más amplia del comportamiento del usuario frente a los asistentes digitales y contribuirá al diseño de tecnologías más eficientes, aceptadas y alineadas con las necesidades de las personas (Marriott & Pitardi, 2021; Van Doorn et al., 2016).

Referencias

Aggarwal, P., & McGill, A. L. (2007). Is That Car Smiling at Me? Schema Congruity as a Basis for Evaluating Anthropomorphized Products. *Journal of Consumer Research*, *34*(4), 468-479. https://doi.org/10.1086/518544

Ajzen, I., & Fishbein, M. (1972). Attitudes and normative beliefs as factors influencing behavioral intentions. *Journal of Personality and Social Psychology*, 21, 1-9. https://doi.org/10.1037/h0031930

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice-Hall.

Balakrishnan, J., & Dwivedi, Y. K. (2021). Conversational commerce: Entering the next stage of AI-powered digital assistants. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04049-5

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, *1*(1), 71-81. https://doi.org/10.1007/s12369-008-0001-3

Basak, E., Gumussoy, C. A., & Calisir, F. (2015). Examining the Factors Affecting PDA Acceptance among Physicians: An Extended Technology Acceptance Model. *Journal of Healthcare Engineering*, 6, 582071. https://doi.org/10.1260/2040-2295.6.3.399

Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2020). Service robot implementation: a theoretical framework and research agenda. T*he service industries journal*, *40*(3-4), 203-225.

Brill, T. M., Munoz, L., & Miller, R. J. (2019). Siri, Alexa, and other digital assistants: A study of customer satisfaction with artificial intelligence applications. *Journal of Marketing Management*, *35*(15-16), 1401-1436. https://doi.org/10.1080/0267257X.2019.1687571

Cao, G., Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2021). Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. *Technovation*, 106, 102312. https://doi.org/10.1016/j.technovation.2021.102312

Chuah, S. H. W., & Yu, J. (2021). The future of service: The power of emotion in human-robot interaction. *Journal of Retailing and Consumer Services*, 61, 102551.

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539-548. https://doi.org/10.1016/j.future.2018.01.055

Cortina, K. G., Sánchez, Y., & De la Garza, M. H. (2024). Efecto de los atributos humanos de los asistentes digitales en la intención de su uso. *RIIIT Revista Internacional de Investigación e Innovación Tecnológica*, *12*(70), 105-123.

Davis, F. D. (1989). Utilidad percibida, facilidad de uso percibida y aceptación del usuario de la tecnología de la información. *MIS Quarterly* , *13*(3), 319-340.

De Cicco, R., Silva, S. C., & Alparone, F. R. (2020). Millennials' attitude toward chatbots: An experimental study in a social relationship perspective. *International Journal of Retail y Distribution Management*, *48*(11), 1213-1233. https://doi.org/10.1108/IJRDM-12-2019-0406

Ferreira, I. (2021). *Estudio de las experiencias del consumidor con tecnologías inteligentes controladas por voz* [Universidad de Zaragoza]. https://dialnet.unirioja.es/servlet/tesis?codigo=304874

Griol, D., & Callejas, Z. (2016). Mobile Conversational Agents for Context-Aware Care Applications. *Cognitive Computation*, *8*(2), 336-356. https://doi.org/10.1007/s12559-015-9352-x

Han, S., & Yang, H. (2018). Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. *Industrial Management and Data Systems*, *118*(3), 618-636. https://doi.org/10.1108/IMDS-05-2017-0214

Huang, J., Wang, L., & Chan, E. (2024). When does anthropomorphism hurt? How tool anthropomorphism negatively affects consumers' rewards for tool users. *Journal of Business Research*, 170, 114355. https://doi.org/10.1016/j.jbusres.2023.114355

IMMUNE, T. I. (2023). *7 ejemplos de uso de inteligencia artificial en el día a día*. https://immune.institute/blog/7-ejemplos-de-uso-de-inteligencia-artificial-en-nuestro-dia-a-dia

Jain, R., Garg, N., & Khera, S. N. (2022). Adoption of AI-Enabled Tools in Social Development Organizations in India: An Extension of UTAUT Model. *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.893691

Knight, W. (2012, mayo 29). *Where Speech Recognition Is Going*. MIT Technology Review. https://www.technologyreview.com/2012/05/29/255893/where-speech-recognition-is-going/

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2, 1-31. https://doi.org/10.1145/3274371

Li, L., Lee, K. Y., Emokpae, E., & Yang, S.-B. (2021). What makes you continuously use chatbot services? Evidence from chinese online travel agencies. *Electronic Markets*, *31*(3), 575-599. https://doi.org/10.1007/s12525-020-00454-z

Lubbe, I., & Ngoma, N. (2021). Useful chatbot experience provides technological satisfaction: An emerging market perspective. *South African Journal of Information Management*. https://sajim.co.za/index.php/sajim/article/view/1299/1953

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46-60. https://doi.org/10.1016/j.futures.2017.03.006

Malhotra, G., & Ramalingam, M. (2023). Perceived anthropomorphism and purchase intention using artificial intelligence technology: Examining the moderated effect of trust. *Journal of Enterprise Information Management*. https://doi.org/10.1108/JEIM-09-2022-0316

Marikyan, D., Papagiannidis, S., Rana, O. F., Ranjan, R., & Morgan, G. (2022). "Alexa, let's talk about my productivity": The impact of digital assistants on work productivity. *Journal of Business Research*, 142, 572-584. https://doi.org/10.1016/j.jbusres.2022.01.015

Marriott, H., & Pitardi, V. (2021). Alexa, she's not human but... Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology and Marketing*, 38. https://doi.org/10.1002/mar.21457

McLean, G., Osei-Frimpong, K., & Barhorst, J. (2021). Alexa, do voice assistants influence consumer brand engagement? – Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of Business Research*, 124, 312-328. https://doi.org/10.1016/j.jbusres.2020.11.045

Medina, J. M., Ábrego, D., & Echeverría-Ríos, O. (2021). Satisfacción, facilidad de uso y confianza del ciudadano en el gobierno electrónico. *Investigación Administrativa*, *50*(127), 23-41.

Moriuchi, E. (2021). An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior. *Psychology* and *Marketing*, *38*(1), 21-42.

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics y Automation Magazine*, 19, 98-100. https://doi.org/10.1109/MRA.2012.2192811

Moussawi, S., & Koufaris, M. (2019). Perceived Intelligence and Perceived Anthropomorphism of Personal Intelligent Agents: Scale Development and Validation. http://hdl.handle.net/10125/59452

Pantano, E., & Pizzi, G. (2020). Forecasting artificial intelligence on online customer assistance: Evidence from chatbot patents analysis. *Journal of Retailing and Consumer Services*, 55, 102096. https://doi.org/10.1016/j.jretconser.2020.102096

Ramadan, Z. B. (2021). "Alexafying" shoppers: The examination of Amazon's captive relationship strategy. *Journal of Retailing and Consumer Services*, 62, 102610. https://doi.org/10.1016/j.jretconser.2021.102610

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places* (pp. xiv, 305). Cambridge University Press.

Roy, R., Babakerkhell, M. D., Mukherjee, S., Pal, D., & Funilkul, S. (2022). Evaluating the Intention for the Adoption of Artificial Intelligence-Based Robots in the University to Educate the Students. *IEEE Access*, 10, 125666-125678. https://doi.org/10.1109/ACCESS.2022.3225555

Sáiz, M. C., Marticorena-Sánchez, R., & Ochoa-Orihuel, J. (2020). Effectiveness of Using Voice Assistants in Learning: A Study at the Time of COVID-19. *International Journal of Environmental Research and Public Health*, *17*(15), Article 15. https://doi.org/10.3390/ijerph17155618

Sharma, M., Joshi, S., Luthra, S., & Kumar, A. (2022). Impact of Digital Assistant Attributes on Millennials' Purchasing Intentions: A Multi-Group Analysis using PLS-SEM, Artificial Neural Network and fsQCA. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10339-5

Sharples, M. (1996). Human-Computer Interaction. *Artificial Intelligence Handbook of Perception and Cognition*, 293-323. https://doi.org/10.1016/B978-012161964-0/50012-1

Silva, F. A., Shojaei, A. S., & Barbosa, B. (2023). Chatbot-Based Services: A Study on Customers' Reuse Intention. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(1), Article 1. https://doi.org/10.3390/jtaer18010024

Song, S. W., & Shin, M. (2024). Uncanny Valley Effects on Chatbot Trust, Purchase Intention, and Adoption Intention in the Context of E-Commerce: The Moderating Role of Avatar Familiarity. *International Journal of Human–Computer Interaction*, *40*(2), 441-456. https://doi.org/10.1080/10447318.2022.2121038

Thormundsson, B. (2023, julio 17). *Artificial Intelligence market size 2030*. Statista. https://www.statista.com/statistics/1365145/artificial-intelligence-market-size/

Van Doorn, J., Mende, M., Noble, S., Hulland, J., Ostrom, A., Grewal, D., & Petersen, J. (2016). Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in Organizational Frontlines and Customers Service Experiences. *Journal of Service Research*, 20. https://doi.org/10.1177/1094670516679272

Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27, 425-478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157-178. https://doi.org/10.2307/41410412

Wang, C., Fayaz Ahmad, S., Ayassrah, A., Awwad, E., Irshad, M., Ali, Y., Al-Razgan, M., & Khan, Y. (2023). An empirical evaluation of technology acceptance model for Artificial Intelligence in E-commerce. *Heliyon*, 9, e18349.

Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1259531

26.

Del andamiaje con inteligencia artificial a la autonomía: evaluación de código estadístico en ingeniería

Francisco Antonio Torres-Espriú
Instituto Tecnológico de Sonora, México
ORCID: 0000-0002-8403-742X

Itzia Nallely Guzmán Mejía
Instituto Tecnológico de Sonora, México
ORCID: 0009-0006-3576-9823

Francisco Guadalupe Avendaño Esparza
Instituto Tecnológico de Sonora, México
ORCID: 0009-0009-4031-8463

Mario Alberto Domínguez-Rovira
Instituto Tecnológico de Sonora, México
ORCID: 0000-0001-6670-2468

DOI del capítulo del libro: https://doi.org/10.56162/transdigitalbc12.26

26.

Del andamiaje con inteligencia artificial a la autonomía: evaluación de código estadístico en ingeniería

Introducción

La educación en ingeniería atraviesa un cambio de paradigma impulsado por la democratización de la inteligencia artificial (IA). En particular, la enseñanza de la programación y el análisis de datos ha pasado de un enfoque puramente sintáctico a uno asistido, donde herramientas de IA generativa (IAG) actúan como colaboradores en la escritura de código. Sin embargo, esta integración no está exenta de riesgos pedagógicos. El presente trabajo nace de una observación crítica en las aulas de una universidad pública del sur de Sonora, México: la creciente dependencia de los estudiantes hacia los *Large Language Models* (LLM) y la consecuente erosión de su autonomía cognitiva.

Bajo el marco propuesto por Ouyang y Jiao (2021), la relación entre el estudiante y la tecnología ha transitado hacia un paradigma de *aprendiz como colaborador*, donde la IA no solo dirige el aprendizaje, sino que apoya la ejecución de tareas complejas. Este modelo se alinea con el concepto clásico de andamiaje de Vygotsky (1978), que sugiere que el aprendizaje se facilita mediante apoyos temporales. En el contexto de la estadística descriptiva, el lenguaje *Python* representa una barrera de entrada significativa. Como señalan Thorgeirsson y Weidmann (2024), la programación textual impone una carga extrínseca que a menudo compite con los recursos mentales necesarios para comprender los conceptos estadísticos. La IA, en teoría, debería reducir esta carga al automatizar la sintaxis, permitiendo que el alumno se concentre en la semántica de los datos.

No obstante, la realidad observada tras ocho semanas de instrucción sugiere un fenómeno de sobredependencia. Durante este periodo, la IA no fue una imposición docente, sino un recurso voluntario para la depuración de errores (*debugging*) y la clarificación sintáctica. A pesar de esta libertad, se detectó una *ilusión de competencia*: al delegar la lógica de programación a la herramienta, el estudiante percibe una maestría que se desvanece

al retirar el asistente. Esta investigación documenta un experimento pedagógico donde se eliminó el andamiaje digital en una evaluación crítica. El objetivo no fue medir el rendimiento académico como un fin aislado, sino analizar cómo la asistencia constante puede, paradójicamente, inhibir la formación de esquemas mentales autónomos.

Fundamentación teórica

Para comprender el impacto de la IAG en la enseñanza de la programación estadística, es imperativo analizar la intersección entre la didáctica de la estadística, las demandas cognitivas de los lenguajes de programación y los nuevos paradigmas de interacción humano-computadora.

La evolución de la enseñanza estadística y el reto de la programación

La educación estadística en el nivel superior ha transitado de un enfoque centrado en el cálculo manual de fórmulas a uno orientado al Análisis Exploratorio de Datos (EDA) y el pensamiento computacional. Autores como Tishkovskaya y Lancaster (2012) señalan que el siglo XXI exige que los estudiantes no solo comprendan la probabilidad teórica, sino que sean capaces de manipular, visualizar e interpretar conjuntos de datos reales. En este contexto, el uso de herramientas tecnológicas dejó de ser opcional para convertirse en el núcleo de la alfabetización estadística (Ben-Zvi, 2000).

Recientemente, lenguajes de programación como *Python* se han consolidado como el estándar para la ciencia de datos debido a la versatilidad de librerías como *Pandas* y *Matplotlib*. Riyantoko et al. (2025) destacan que el autoaprendizaje de la estadística fundamental mediante *Python* permite a los estudiantes implementar soluciones de nivel profesional. Sin embargo, la transición hacia el uso de código textual en etapas tempranas de formación introduce barreras pedagógicas severas. A diferencia de las interfaces gráficas de usuario (GUI) tradicionales, la programación requiere una precisión absoluta; la omisión de un solo carácter (un error de sintaxis) impide la ejecución del análisis, generando frustración y desviando la atención del estudiante: del concepto estadístico al error informático.

Teoría de la carga cognitiva en la programación textual

El aprendizaje simultáneo de la estadística y la programación puede analizarse a través de la lente de la teoría de la carga cognitiva propuesta por Sweller (1988). Esta teoría postula que la memoria de trabajo humana es limitada y que la instrucción debe diseñarse

para optimizar estos recursos. Sweller divide la carga cognitiva en tres dimensiones: intrínseca (la dificultad inherente a la tarea, como entender la Regla de Sturges), extrínseca (la carga innecesaria generada por el formato de instrucción o las herramientas) y germana (el esfuerzo mental constructivo que crea esquemas en la memoria a largo plazo).

En la educación en ingeniería en software, los programadores novatos enfrentan una carga extrínseca desproporcionada. Robins et al. (2003) documentaron que los principiantes carecen de los esquemas mentales que permiten a los expertos leer el código de forma fluida. Más recientemente, Thorgeirsson y Weidmann (2024) confirmaron empíricamente que *Python* impone una carga cognitiva intrínseca y extrínseca significativamente mayor en comparación con lenguajes de programación visual. Cuando un estudiante intenta calcular una distribución de frecuencias en *Python*, su memoria de trabajo se satura tratando de recordar si el comando correcto es value_counts(), cut() o hist(), lo que deja recursos mentales insuficientes (carga germana) para interpretar qué significa realmente esa distribución en el contexto del problema.

El andamiaje cognitivo y los paradigmas de la inteligencia artificial

Para mitigar esta saturación cognitiva, la educación ha recurrido históricamente al concepto *andamiaje* (*scaffolding*), teorizado por Vygotsky (1978). El andamiaje consiste en proporcionar apoyos temporales que permiten al estudiante resolver problemas que están más allá de su capacidad de desarrollo independiente (su zona de desarrollo próximo), con el objetivo de retirar gradualmente dicha ayuda a medida que se adquiere competencia.

Con la democratización de modelos como *ChatGPT* o *Google Gemini*, el andamiaje ha evolucionado de ser una intervención docente a una asistencia algorítmica constante. Ouyang y Jiao (2021) proponen que la integración de la IA en la educación se divide en tres paradigmas: IA-dirigida (el alumno como receptor), IA-soportada (el alumno como colaborador) e IA-empoderada (el alumno como líder). El uso de la IAG para la depuración de errores (*debugging*) y la generación de bloques de código se sitúa en el segundo paradigma. En este escenario de colaboración, la herramienta asume la carga extrínseca (la sintaxis) y teóricamente libera al estudiante para que asuma el rol de director conceptual. La IA actúa como un copiloto que traduce la intención estadística del estudiante a instrucciones legibles por la máquina.

Riesgos de la asistencia generativa: superficialidad y dependencia

A pesar de los beneficios evidentes en la fluidez del trabajo, la asistencia generativa introduce riesgos críticos para la consolidación del aprendizaje. Baidoo-Anu y Ansah (2023) advierten que, si bien estas herramientas promueven un aprendizaje interactivo y reducen el estancamiento técnico, su uso no supervisado puede fomentar una *superficialidad cognitiva*.

El problema radica en la naturaleza de la fricción cognitiva. Un grado manejable de frustración y resolución manual de errores es necesario para que el cerebro forje rutas de memoria a largo plazo. Al delegar sistemáticamente la resolución de conflictos sintácticos a la IA, los estudiantes experimentan la ilusión de competencia técnica; producen cuadernos interactivos funcionales y visualizaciones complejas, pero sin interiorizar la lógica del código subyacente. La literatura advierte que este fenómeno de sobredependencia se desenmascara abruptamente cuando el andamiaje digital es retirado, momento en el cual el estudiante se revela incapaz de replicar de forma autónoma operaciones que ejecutaba con éxito bajo asistencia, evidenciando una brecha sustancial entre el desempeño asistido y el aprendizaje real.

Metacognición y el problema de la transferencia en entornos asistidos

Un factor determinante para el éxito de cualquier andamiaje educativo es la metacognición: la capacidad del estudiante para monitorear y evaluar su propio nivel de comprensión. En los entornos de aprendizaje tradicionales, los errores de compilación o ejecución actúan como retroalimentación inmediata, obligando al alumno a reflexionar sobre su lógica algorítmica. Sin embargo, en un entorno asistido por IA, esta retroalimentación se altera profundamente.

Cuando un LLM como *Google Gemini* corrige un error de sintaxis o sugiere el uso de la función estadística correcta, el estudiante a menudo omite el proceso de reflexión analítica. Al respecto, la literatura sobre autorregulación del aprendizaje sugiere que la fluidez con la que se resuelven los problemas técnicos a través de prompts (instrucciones en lenguaje natural) genera un sesgo de atribución. Los estudiantes experimentan lo que se conoce como *ilusión de competencia*: confunden el rendimiento del sistema sociotécnico (Alumno + IA) con su propia competencia individual.

Este fenómeno se vuelve crítico al analizar el concepto de *transferencia del aprendizaje* (la capacidad de aplicar lo aprendido en un contexto a una situación nueva). La verdadera

apropiación del conocimiento en ingeniería en software exige que las habilidades desarrolladas en un entorno de práctica (con andamiaje) sean transferibles a un entorno de ejecución pura (sin andamiaje). Si la IAG asume la carga cognitiva de manera permanente en lugar de actuar como un soporte temporal en la zona de desarrollo próximo, el estudiante no construye los esquemas de memoria a largo plazo. En consecuencia, al enfrentar escenarios de desconexión digital, la falta de transferencia se hace evidente, resultando en una parálisis operativa ante tareas básicas de importación de datos o aplicación de fórmulas estadísticas.

Método de investigación

Participantes y contexto

La presente investigación empírica se llevó a cabo bajo un enfoque cuantitativo de alcance descriptivo e inferencial. La muestra estuvo conformada por 27 estudiantes (n = 27) inscritos en el programa de Ingeniería en Software de una universidad pública ubicada en el sur de Sonora, México. Los participantes cursaban la asignatura de probabilidad y estadística, ubicada en el tercer semestre de su programa educativo. El grupo poseía conocimientos previos en el lenguaje *Python*, adquiridos durante la asignatura de Programación I en su primer semestre; sin embargo, se encontraban en su primer acercamiento formal a las librerías especializadas en ciencia de datos (*Pandas*, *Matplotlib*).

Procedimiento y fase de instrucción asistida

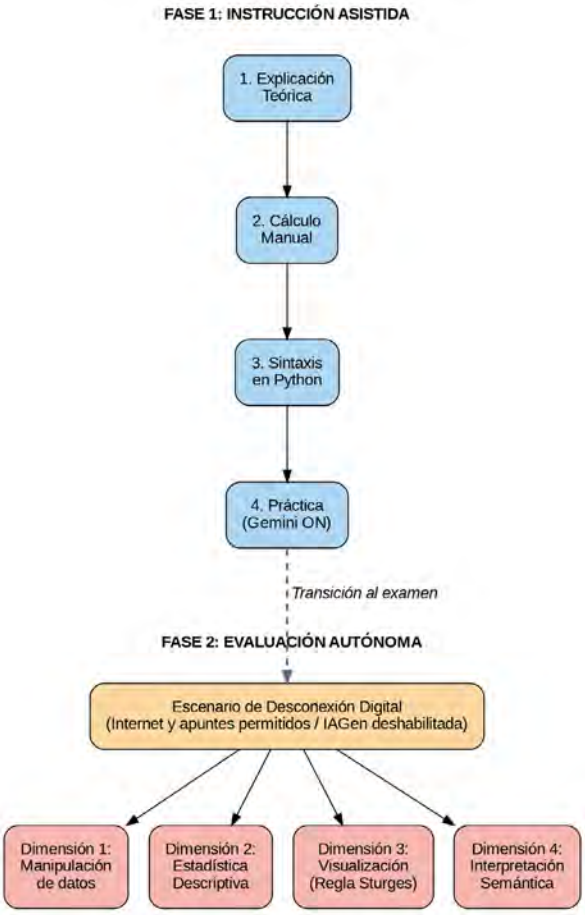
El diseño del estudio se estructuró en dos fases principales: un periodo de instrucción asistida y una evaluación de autonomía (Figura 1). La primera fase abarcó ocho semanas de clases teórico-prácticas. Durante este periodo, el docente instruyó a los estudiantes en técnicas de EDA, abarcando desde la importación de bases de datos hasta el cálculo de medidas de tendencia central, dispersión y la generación de gráficos estadísticos.

A lo largo de estas ocho semanas, la dinámica pedagógica habitual siguió una secuencia de cuatro pasos: 1) explicación teórica del concepto matemático, 2) resolución de un ejercicio mediante cálculo manual, 3) explicación docente de la sintaxis del código correspondiente, y 4) prácticas de ejecución en cuadernos interactivos. Durante esta última etapa práctica, se implementó un modelo de andamiaje digital abierto (*scaffolding*). Los estudiantes tuvieron libertad para utilizar *Google Gemini* como herramienta de soporte concurrente para la

depuración de errores (*debugging*) y la consulta rápida de sintaxis, interviniendo cuando se presentaban bloqueos técnicos.

Figura 1

Diseño metodológico: De la instrucción asistida a la evaluación autónoma



Instrumento de recolección de datos

Para la segunda fase, se diseñó un instrumento de evaluación práctica estructurado en un cuaderno interactivo. El examen utilizó un conjunto de datos (*dataset*) real que contenía las especificaciones técnicas y precios de diversas computadoras portátiles. El instrumento constó de 16 tareas programáticas y analíticas secuenciales que evaluaron cuatro dimensiones estadísticas:

1. Manipulación de datos: Importación de librerías, lectura de archivos CSV e identificación (sin alteración ni limpieza) de valores nulos.

2. Estadística descriptiva: Cálculo de promedios, medianas, desviaciones estándar y varianzas sobre las características de los equipos.

3. Visualización de datos: Construcción de histogramas y gráficos de dispersión. Un elemento crítico fue la obligatoriedad de calcular manualmente el número de clases mediante la Regla de Sturges para los histogramas.

4. Interpretación y argumentación: Redacción de explicaciones sobre el significado de los resultados obtenidos (ej. la asimetría de la distribución o el comportamiento de los datos), requiriendo un análisis semántico más allá de la mera ejecución del código.

Escenario de evaluación y análisis de datos

La evaluación se aplicó en un entorno presencial estrictamente controlado, denominado escenario de desconexión algorítmica. A diferencia de una evaluación tradicional cerrada, no se restringió el acceso a internet ni la consulta de los cuadernos de práctica elaborados durante las ocho semanas previas. La condicionante estricta fue la inhabilitación del asistente *Google Gemini* dentro de la interfaz de *Google Colab* y la prohibición de utilizar herramientas de IAG alternativas.

La recolección de datos se efectuó mediante la asignación de puntajes a las 16 tareas del entorno de ejecución en *Python* (evaluación de desconexión). Adicionalmente, para evaluar el impacto del andamiaje algorítmico previo, se cuantificó el porcentaje de cumplimiento de las prácticas formativas desarrolladas por cada estudiante durante las ocho semanas de instrucción asistida por IAG. Los datos recolectados fueron exportados a un archivo estructurado para su posterior análisis estadístico, buscando correlacionar el nivel de trabajo asistido con la capacidad real de ejecución autónoma e identificar las dimensiones cognitivas donde se originó la mayor parálisis operativa.

Resultados

El análisis de los datos recolectados se estructuró en tres fases: primero, una evaluación descriptiva del desempeño independiente; segundo, un análisis de correlación entre el trabajo asistido previo y el éxito autónomo; y tercero, la identificación de los cuellos de botella técnicos en la escritura de código.

Análisis del desempeño en el escenario de desconexión

El primer objetivo consistió en evaluar la capacidad de los estudiantes (N = 27) para resolver problemas de estadística descriptiva en *Python* sin asistencia externa. Los estadísticos descriptivos (Tabla 1) muestran una media grupal de 67.52 puntos (de 100 posibles), con una dispersión amplia (DE = 24.27). Esta variabilidad indica una marcada heterogeneidad en el rendimiento: mientras un subgrupo de estudiantes logró mantener la funcionalidad del código, otro sector evidenció un déficit significativo en su capacidad de ejecución al retirar el acceso a la IAG, registrando puntajes mínimos de 10.0.

Tabla 1

Estadísticos descriptivos del desempeño de los estudiantes

Estadístico	Examen Python (Autónomo)
N	27
Media	67.52
Desviación Estándar (DE)	24.27
Mínimo	10.00
Máximo	96.00
Shapiro-Wilk (p)	0.014*

Nota. * p < 0.05 indica desviación de la normalidad.

La prueba de Shapiro-Wilk confirmó que la distribución de las calificaciones se desvía de la normalidad (p = 0.014). Este hallazgo empírico no constituye una anomalía en los datos, sino que evidencia la brecha de desempeño generada por la ausencia de la IA; el retiro del soporte tecnológico segmentó a la muestra, distanciando a quienes asimilaron la lógica de programación de quienes dependían del asistente para la depuración de errores. Esta dispersión se visualiza en la Figura 2.

Correlación entre el trabajo asistido y la autonomía técnica

Se analizó si el cumplimiento de las prácticas formativas (donde se permitió el uso de IAGen) actuaba como un predictor del éxito en la evaluación autónoma. Dada la asimetría de la distribución, se utilizó el coeficiente de correlación de Spearman (ρ). Los resultados se presentan en la Tabla 2 y se proyectan en la Figura 3.

Figura 2

Distribución de calificaciones en desconexión (Python)

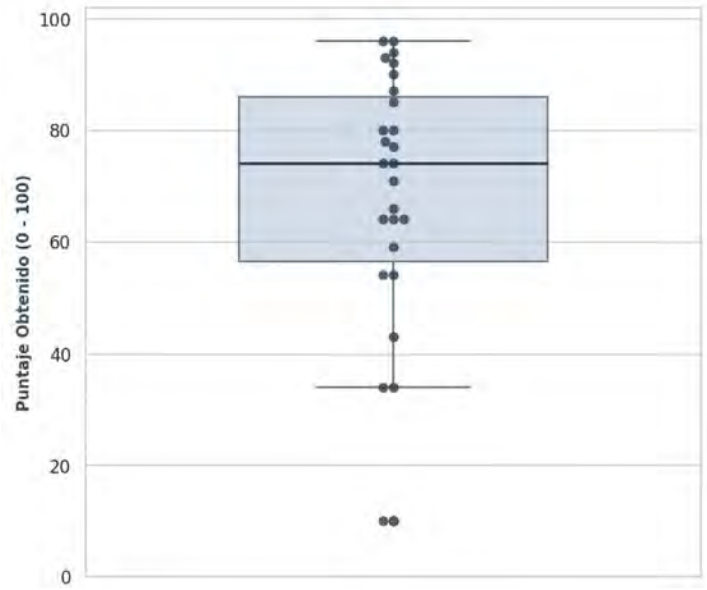


Tabla 2

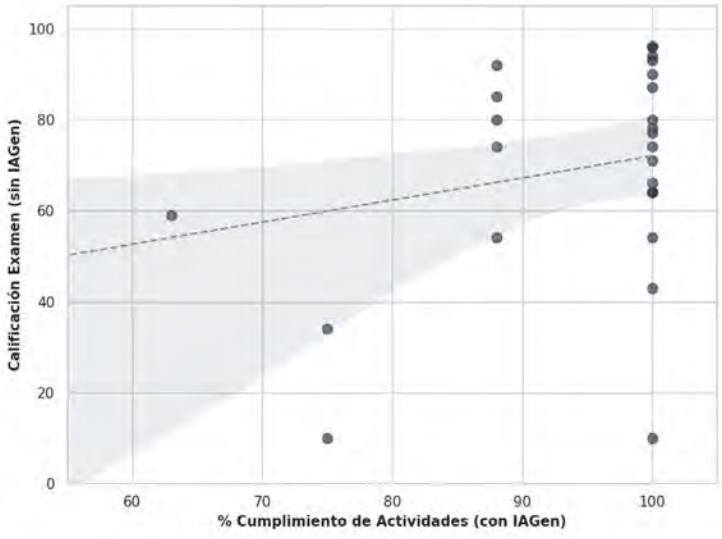
Correlación de Spearman entre cumplimiento de actividades (con IA) y examen autónomo (sin IA)

Variable	ρ de Spearman	Valor *p*
% Cumplimiento vs. Puntaje Examen Python	0.354	0.070

Nota. * La correlación no es estadísticamente significativa al nivel de α = 0.05.

Figura 3

Correlación entre práctica asistida y desempeño autónomo



El análisis arrojó una correlación positiva débil (ρ = 0.354) que no alcanzó significancia estadística (p = .070). Este comportamiento sugiere que el progreso constante en las tareas asistidas no se tradujo sistemáticamente en la adquisición de competencias individuales. La fluidez obtenida mediante el andamiaje algorítmico parece haber enmascarado una carencia de aprendizaje profundo, la cual emergió al requerirse una ejecución técnica independiente.

Evaluación de dimensiones técnicas y lógica de programación

Al examinar los cuadernos de programación, se identificaron patrones de error recurrentes que permiten categorizar las barreras operativas de los estudiantes:

1. Configuración e inicialización: La mayoría de los participantes mostró un dominio consolidado en tareas rutinarias, como la importación de librerías (*import pandas as pd*). Estos comandos, al poseer una estructura altamente repetitiva, lograron integrarse en la memoria a largo plazo.

2. Lógica funcional y sintáctica: Se detectaron dificultades al invocar métodos específicos de estadística descriptiva. Los estudiantes demostraron comprensión del concepto estadístico, pero presentaron bloqueos cognitivos al intentar evocar la sintaxis precisa del lenguaje sin las sugerencias de autocompletado de la IA.

3. Integración de parámetros y Regla de Sturges: Esta dimensión representó el mayor obstáculo técnico. Aunque se solicitó calcular el número de intervalos mediante la Regla de Sturges ($k = 1 + 3.322 \log n$), una proporción considerable de los estudiantes fue incapaz de integrar este valor en el código. Generaron histogramas omitiendo el parámetro solicitado y utilizando la configuración por defecto de *Matplotlib*. Este patrón sugiere que la IA operaba como el traductor principal entre la teoría matemática y la implementación técnica; ante su ausencia, el estudiante optó por soluciones automatizadas básicas, ignorando las restricciones explícitas del problema.

4. Carga cognitiva e interpretación: Se observó que el esfuerzo mental invertido en la resolución manual de errores sintácticos limitó la capacidad de los estudiantes para redactar interpretaciones precisas. La saturación de la memoria de trabajo por la exigencia del entorno de programación redujo la profundidad del análisis estadístico, demostrando que la fluidez sintáctica es un requisito previo para la argumentación analítica.

Discusión

Los resultados obtenidos invitan a una reflexión profunda sobre la integración de la IAG en el currículo de ingeniería. Lejos de cuestionar la eficacia de la tecnología, los hallazgos subrayan la necesidad de replantear las estrategias didácticas para asegurar que la asistencia algorítmica actúe como un facilitador del aprendizaje, y no como un sustituto que eclipse el desarrollo de habilidades cognitivas fundamentales.

La disparidad entre el trabajo formativo y el desempeño autónomo

La ausencia de una correlación estadísticamente significativa ($\rho = 0.354$; $p = 0.070$) entre el cumplimiento de las actividades durante las ocho semanas de instrucción y el resultado de la evaluación individual constituye un hallazgo central. Si bien la literatura reciente (Riyantoko et al., 2025) enfatiza las bondades de *Python* y la IA para fomentar el autoaprendizaje, los datos de la presente investigación sugieren que el éxito sostenido en la entrega de tareas no siempre es un indicador fiable de la apropiación del conocimiento subyacente.

Esta brecha de desempeño es explicable a través del fenómeno de la *ilusión de competencia* (Baidoo-Anu y Ansah, 2023). Al contar con la posibilidad voluntaria de recurrir a herramientas como *Google Gemini* para la clarificación sintáctica o la resolución de errores durante sus prácticas, es factible que el estudiante experimente una percepción de dominio técnico elevada. Sin embargo, al retirar este andamiaje en un entorno controlado, se evidencia que dicho aprendizaje no logró consolidarse en esquemas mentales independientes (Vygotsky, 1978). El desafío pedagógico, por lo tanto, no reside en la herramienta tecnológica per se, sino en la dificultad del estudiante para autorregular su uso y efectuar una transición efectiva desde la asistencia externa hacia la autonomía técnica.

Carga cognitiva y la gestión de restricciones técnicas

La Teoría de la Carga Cognitiva (Sweller, 1988) proporciona un marco explicativo idóneo para comprender el decaimiento del desempeño ante tareas que exigían precisión matemática y estructural, como la aplicación de la Regla de Sturges. Durante la evaluación de desconexión, los alumnos se enfrentaron a una carga extrínseca considerable al tener que gestionar y evocar la sintaxis estricta de librerías como *Matplotlib* sin soporte concurrente.

El hecho de que una proporción significativa de la muestra optara por utilizar los parámetros por defecto de las funciones de visualización, omitiendo el cálculo algorítmico

solicitado, apunta hacia una estrategia de *ahorro cognitivo*. Ante la saturación de la memoria de trabajo por las demandas informáticas del lenguaje de programación, el estudiante tiende a sacrificar la precisión analítica para asegurar una compilación funcional mínima. Como argumentan Thorgeirsson y Weidmann (2024), la IA suele operar como un puente invisible que resuelve la fricción entre la teoría estadística y el código textual; si el alumno no ejercita manualmente la construcción de ese puente, pierde la capacidad de orquestar el análisis de datos bajo presión operativa.

Hacia un modelo de andamiaje digital responsable

La evidencia recolectada no sugiere que el uso de la IAG en el aula sea contraproducente, sino que su adopción bajo el paradigma del *alumno como colaborador* (Ouyang y Jiao, 2021) requiere una calibración cuidadosa durante las etapas tempranas de formación en ingeniería. La IA representa un recurso invaluable para mitigar la frustración inicial frente a la programación, pero si el entorno educativo permite que esta asuma permanentemente el proceso de depuración (*debugging*), se corre el riesgo de formar profesionales con una resiliencia técnica limitada.

Para capitalizar el potencial de la IA como aliado educativo, se sugiere que la práctica docente incorpore mecanismos de andamiaje transitorio (*scaffolding with fading*). Esto implica diseñar momentos de *desconexión programada* a lo largo del curso, donde el estudiante confronte sus propias limitaciones sin la red de seguridad del asistente digital. El objetivo último de la educación en software no es erradicar la IA del flujo de trabajo, sino empoderar al estudiante para que asuma el rol directivo, garantizando que la solidez conceptual preceda y guíe a la automatización algorítmica.

Conclusiones

El presente estudio aporta evidencia empírica sobre los efectos de la IAG en el aprendizaje de la programación aplicada a la estadística en estudiantes de ingeniería en software. Se concluye que el uso continuo y no regulado de estas herramientas fomenta una *ilusión de competencia*. Aunque los estudiantes logran resolver y entregar productos de alta complejidad bajo asistencia algorítmica, la remoción del soporte tecnológico revela deficiencias críticas en la asimilación de la sintaxis y en la capacidad de estructurar soluciones de forma autónoma.

Desde la perspectiva de la carga cognitiva, la delegación prematura de la escritura y depuración de código a la IA impide que el estudiante consolide los esquemas mentales necesarios para la programación. El déficit observado en la integración de parámetros matemáticos —como la omisión de la Regla de Sturges— ilustra cómo, ante la falta de dominio técnico, el esfuerzo mental se agota en la operatividad básica, sacrificando el rigor analítico. Por lo tanto, se reafirma que la fluidez sintáctica sigue siendo un prerrequisito ineludible para alcanzar niveles superiores de análisis e interpretación de datos.

Lejos de proponer la restricción o prohibición de la IAG en el entorno universitario, los hallazgos exigen transitar hacia una integración pedagógica estructurada. Es imperativo que las instituciones educativas diseñen estrategias de andamiaje transitorio que incluyan escenarios de "desconexión algorítmica" obligatoria. Estas pausas evaluativas son necesarias para que el estudiante confronte la fricción cognitiva propia del aprendizaje, asegurando que la tecnología funcione como un catalizador del pensamiento crítico y no como un simple atajo procedimental.

Finalmente, es pertinente reconocer las limitaciones de esta investigación, particularmente el tamaño de la muestra y la falta de un mecanismo de trazabilidad directa sobre el nivel exacto de interacción que cada estudiante mantuvo con *Google Gemini* durante el trabajo extraclase. Futuras líneas de investigación deberán contemplar diseños cuasi-experimentales con grupos de control, así como análisis cualitativos de la ingeniería de instrucciones (*prompt engineering*) empleada por los alumnos. Esto permitirá comprender con mayor profundidad cómo la naturaleza de la interacción humano-computadora moldea la adquisición de competencias definitivas en la ingeniería moderna.

Referencias

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52-62. https://doi.org/10.2139/ssrn.4337484

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning*, *2*(1-2), 127-155. https://doi.org/10.1207/S15327833MTL0202_6

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Riyantoko, P. A., Funabiki, N., Brata, K. C., Mentari, M., Damaliana, A. T., & Kaswar, A. B. (2025). A fundamental statistics self-learning method with Python programming for data science implementations. *Information*, *16*(7), 607. https://doi.org/10.3390/info16070607

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, *13*(2), 137-172. https://doi.org/10.1076/csed.13.2.137.14200

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Thorgeirsson, S., Weidmann, T. B., Weidmann, K.-H., & Su, Z. (2024). Comparing cognitive load among undergraduate students programming in Python and the visual language Algot. En *Proceedings of the 55th ACM Technical Symposium on Computer Science Education* (pp. 1328-1334). Association for Computing Machinery. https://doi.org/10.1145/3626252.3630808

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, *20*(2). https://doi.org/10.1080/10691898.2012.11889641

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

INTELIGENCIA ARTIFICIAL

EXPERIENCIAS Y REFLEXIONES SOBRE LA INVESTIGACIÓN CIENTÍFICA

ISBN: 978-968-9724-25-4

Trans digital editorial