

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

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 Attribution 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 Chiatichoua	
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 Eliphaz 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, Lucía 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-Díaz, 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	



04.

**JUSTICIA ALGORÍTMICA Y GOBERNANZA ÉTICA
ANTE LOS SEGOS 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

04.

JUSTICIA ALGORÍTMICA Y GOBERNANZA ÉTICA ANTE LOS SEGOS 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	<i>Scopus, Web of Science, ERIC, SpringerLink, RedALyC, SciELO, UNESCO, OECD, European Commission</i>	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 citas	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 citas. 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 mutandis*, 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.

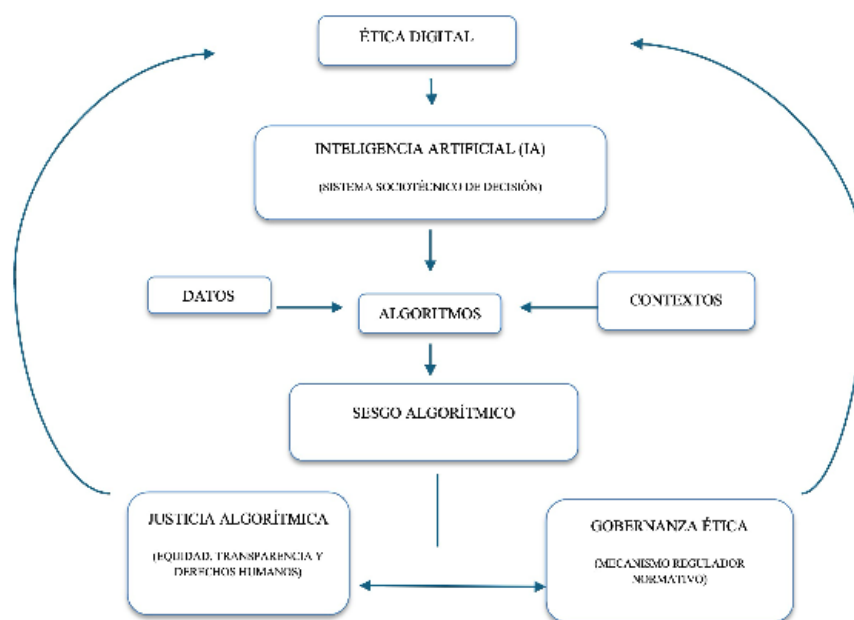
Por tanto, se considera que la IA es sociotécnica: su desempeño y sus efectos dependen tanto de decisiones técnicas como de contextos sociales. El sesgo algorítmico tiene múltiples fuentes y no se resuelve únicamente con más datos; demanda diseño, documentación y evaluación centrados en la equidad y 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 interdisciplinarias (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/rlicts.v17n2.2022>
- Jobin, A., Lenca, 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/ijj.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.



INTELIGENCIA ARTIFICIAL

EXPERIENCIAS Y REFLEXIONES SOBRE LA INVESTIGACIÓN CIENTÍFICA

ISBN: 978-968-9724-25-4



9 789689 724254

Trans[®]
digital
editorial