www.revistatribunal.org

Revista en Ciencias de la Educación y Ciencias Jurídicas

ISSN: 2959-6513 - ISSN-L: 2959-6513 Volumen 5. No. 13 / Octubre – Diciembre 2025
Páginas 243 - 260

Vulnerabilidades en la protección de datos personales de menores ante sistemas de inteligencia artificial: Revisión sistemática

Vulnerabilities in the protection of minors' personal data in the face of artificial intelligence systems: A systematic review

Vulnerabilidades na proteção de dados pessoais de menores face aos sistemas de inteligência artificial: Uma revisão sistemática

Cyntia Raquel Rudas Murga
dra.crrm@outlook.com
Universidad Nacional Mayor de San Marcos. Lima, Perú

Claudio Flores Seefoó
claudiofseefoo@uagro.mx
Universidad Autónoma de Guerrero. Chilpancingo, México

Carlos Iván Ramírez Parra
17478@uagro.mx
Universidad Autónoma de Guerrero. Chilpancingo, México

Adamaris Silva Encarnación
23500063@uagro.mx
Universidad Autónoma de Guerrero. Chilpancingo, México

http://doi.org/10.59659/revistatribunal.v5i13.268

Artículo recibido 7 de agosto 2025 | Aceptado 25 de septiembre 2025 | Publicado 2 de octubre

Palabras clave:
Artificial; Datos personales; Inteligencia; Menores; Sistemas; Vulnerabilidades

Resumen

El avance acelerado de la inteligencia artificial (IA) ha transformado profundamente el ecosistema digital, generando nuevas oportunidades y riesgos en el tratamiento de datos personales. El objetivo del estudio es analizar las vulnerabilidades en la protección de datos personales de menores ante sistemas de inteligencia artificial. El enfoque es cualitativo, descriptivo, b una bajo una revisión sistemática siguiendo directrices PRISMA en el periodo 2020-2025. Las bases de datos consultadas Web of Science, Scopus, PubMed, IEEE Xplore y repositorios especializados. Se identificaron 1,247 estudios iniciales, de los cuales 18 fueron incluidos en la revisión. Los hallazgos revelan prácticas sistemáticas de recopilación inadvertida de datos (89 %) y perfilamiento algorítmico no consentido (78 %), así como exposición a contenidos inapropiados y ciberacoso automatizado. Se concluye que la protección de menores requiere un enfoque integral que implemente principios de privacidad desde el diseño, marcos regulatorios especializados, alfabetización digital crítica en la educación formal y responsabilidad compartida entre industria, reguladores y sociedad civil.

Keywords:

Artificial Intelligence; Personal Data; Intelligence; Minors; Systems; Vulnerabilities

Abstract

The accelerated advancement of artificial intelligence (AI) has profoundly transformed the digital ecosystem, generating new opportunities and risks in the processing of personal data. The objective of this study is to analyze vulnerabilities in the protection of minors' personal data against artificial intelligence systems. The approach is qualitative, descriptive, and based on a systematic review following PRISMA guidelines for the period 2020-2025. The databases consulted were Web of Science, Scopus, PubMed, IEEE Xplore, and specialized repositories. A total of 1,247 initial studies were identified, of which 18 were included in the review. The findings reveal systematic practices of inadvertent data collection (89%) and non-consensual algorithmic profiling (78%), as well as exposure to inappropriate content and automated cyberbullying. It is concluded that the protection of minors requires a comprehensive approach that implements privacy-by-design principles, specialized regulatory frameworks, critical digital literacy in formal education, and shared responsibility between industry, regulators, and civil society.

Palavras-chave:

Inteligência Artificial; Dados Pessoais; Inteligência; Menores; Sistemas; Vulnerabilidades

Resumo

O avanço acelerado da inteligência artificial (IA) transformou profundamente o ecossistema digital, gerando novas oportunidades e riscos no processamento de dados pessoais. O objetivo deste estudo é analisar vulnerabilidades na proteção de dados pessoais de menores contra sistemas de inteligência artificial. A abordagem é qualitativa, descritiva e baseada em uma revisão sistemática seguindo as diretrizes PRISMA para o período de 2020 a 2025. As bases de dados consultadas foram Web of Science, Scopus, PubMed, IEEE Xplore e repositórios especializados. Um total de 1.247 estudos iniciais foram identificados, dos quais 18 foram incluídos na revisão. Os resultados revelam práticas sistemáticas de coleta inadvertida de dados (89%) e criação de perfis algorítmicos não consensuais (78%), bem como exposição a conteúdo inapropriado e cyberbullying automatizado. Conclui-se que a proteção de menores requer uma abordagem abrangente que implemente os princípios de privacidade desde a concepção, marcos regulatórios especializados, alfabetização digital crítica na educação formal e responsabilidade compartilhada entre indústria, órgãos reguladores e sociedade civil.

INTRODUCCIÓN

El avance acelerado de la inteligencia artificial (IA) ha transformado profundamente el ecosistema digital, generando nuevas oportunidades y riesgos en el tratamiento de datos personales. Esta transformación adquiere especial relevancia en el caso de los menores de edad, quienes, por su condición jurídica y cognitiva, constituyen una población especialmente vulnerable frente a sistemas automatizados de recopilación, análisis y generación de información. Según el informe Global Digital 2024, el 95 % de los adolescentes entre 13 y 17 años accede regularmente a internet, y el 87 % utiliza plataformas digitales que integran IA para personalizar contenidos (Hootsuite y We Are Social, 2024).

La emergencia de modelos de IA generativa en 2022 marcó un punto de inflexión en la capacidad de estos sistemas para procesar, analizar y generar contenido basado en datos de usuarios, incrementando exponencialmente los riesgos de privacidad para menores (OpenAI, 2023; García, 2025). Estudios recientes documentan que el 73% de las aplicaciones dirigidas a menores recopilan más datos de los declarados en sus políticas de privacidad (Reyes et al., 2023).

Los diversos sistemas de inteligencia artificial, se integran de manera imperceptible en la vida cotidiana, desde sistemas de recomendación en redes sociales hasta algoritmos en procesos educativos y laborales, al punto de que muchos de sus efectos pasan desapercibidos.

Lo anterior se normaliza en una sociedad altamente industrializada, pues a partir de la cuarta revolución industrial, el objetivo es integrar la tecnología en el quehacer cotidiano, por poner un ejemplo podemos mencionar el internet de las cosas (IOT por sus siglas en inglés), es decir que convivimos con estos sistemas de IA de manera consciente o inconsciente.

El uso de la IA trae innovaciones con usos diversos en la vida cotidiana tanto positivas como negativas, van desde diagnósticos médicos, hasta herramientas de vigilancia masiva. Pero en esas tecnologías subyacen algoritmos con sesgos que pueden representar amenazas que perpetúen y profundizar las desigualdades existentes. Estas amenazas no siempre son evidentes para los usuarios, pero plantean desafíos urgentes en materia de privacidad, equidad y derechos humanos, se habla entonces de amenazas silenciosas.

Estas amenazas no siempre son evidentes para los usuarios, pero plantean desafíos urgentes en materia de privacidad, equidad y derechos humanos. Rada (2025) denomina este fenómeno como 'IA silenciosa', al referirse a sistemas que operan de forma imperceptible y amplifican riesgos sin una evaluación legal rigurosa. En este sentido, la IA generativa plantea desafíos éticos y jurídicos sin precedentes, al permitir la creación de deepfakes y la manipulación de datos biométricos extraídos de redes sociales o entornos educativos.

El marco regulatorio internacional presenta significativos vacíos en la protección específica de datos de menores ante sistemas de IA. Mientras que el Reglamento General de Protección de Datos (RGPD) de la Unión Europea establece principios generales, la mayoría de jurisdicciones carecen de normatividad específica para IA y menores (Kotseva y Tsolova, 2024). En el contexto mexicano, la ausencia de regulación específica sobre IA deja a los menores particularmente vulnerables (Medina y Torres, 2025).

Diversos estudios han documentado que el 73 % de las aplicaciones dirigidas a menores recopilan más datos de los declarados en sus políticas de privacidad, lo que evidencia una brecha entre la práctica tecnológica y la protección legal (Reyes et al., 2023). Esta situación se agrava por la falta de marcos normativos específicos. Aunque el Reglamento General de Protección de Datos (RGPD) de la Unión Europea establece principios generales para el tratamiento de datos personales, no contempla disposiciones concretas sobre IA y menores (Kotseva y Tsolova, 2024). A nivel global, la Organización de Naciones Unidas (ONU) ha instado a establecer una arquitectura inclusiva para la gobernanza de la IA, subrayando la necesidad de proteger los derechos humanos, especialmente de los grupos vulnerables como la infancia (ONU, 2024).

Desde una perspectiva académica, se han identificado esfuerzos parciales para abordar esta problemática. Charisi et al. (2022), desde la Comisión Europea, propusieron una agenda integrada para investigar el impacto de la IA en los derechos de la infancia. Van der Hof et al. (2023) destacaron la necesidad de enfoques interdisciplinarios que integren perspectivas jurídicas, técnicas y sociales, Kumar et al. (2024), subrayaron la urgencia de construir marcos teóricos robustos para comprender las implicaciones de la IA en la privacidad infantil. En Estados Unidos, la Federal Trade Commission (FTC) inició investigaciones sobre el uso de chatbots por menores, alertando sobre riesgos de manipulación emocional y recolección indebida de datos (FTC, 2025).

En el contexto mexicano, estudios como el de Sánchez (2024) han evidenciado la vulnerabilidad jurídica de los menores ante la IA generativa, mientras que Trejo (2024) analizó el interés superior de la niñez en entornos digitales, proponiendo una actualización normativa que contemple los riesgos emergentes. Mendoza Enríquez (2021) identificó desafíos normativos en la protección de datos personales frente a sistemas de IA, señalando la necesidad de mecanismos efectivos para garantizar derechos humanos. A nivel local, el Instituto de Transparencia de Jalisco (ITEI) ha documentado prácticas de recolección de datos sin consentimiento en plataformas educativas, lo que pone de relieve la necesidad de fortalecer la ciberseguridad infantil (ITEI, 2022).

A pesar del creciente interés académico, no existe una revisión sistemática que analice integralmente las vulnerabilidades específicas de los datos personales de menores ante sistemas de IA, considerando tanto aspectos técnicos como marcos normativos y estrategias de protección. En este contexto, el presente estudio tiene como objetivo principal analizar sistemáticamente la literatura sobre las vulnerabilidades en la protección de datos personales de menores ante sistemas de inteligencia artificial, científica publicada entre 2020 y 2025, identificando los principales riesgos, marcos normativos y estrategias de protección documentadas.

Este análisis permitirá responder a las siguientes preguntas de investigación: ¿Cuáles son las principales vulnerabilidades que enfrentan los menores en el tratamiento de sus datos personales por sistemas de IA? ¿Qué marcos normativos existen y cuáles son sus limitaciones? ¿Qué estrategias de protección han sido propuestas y con qué resultados?

La relevancia de esta revisión sistemática de literatura, radica en su contribución a la comprensión integral de un problema emergente que afecta derechos fundamentales de la infancia en entornos digitales. Asimismo, busca aportar insumos para el diseño de políticas públicas y líneas de investigación futuras que promuevan una protección efectiva de los datos personales de menores en la era de la inteligencia artificial.

METODOLOGÍA

Este estudio se desarrolló bajo un enfoque cualitativo, mediante una revisión sistemática de la literatura científica, siguiendo las directrices del protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) propuesto por Page et al. (2021). El objetivo de esta revisión es analizar investigaciones relevantes sobre las vulnerabilidades en la protección de datos personales de menores ante sistemas de inteligencia artificial (IA), publicadas entre enero de 2020 y marzo de 2025.

La estrategia de búsqueda se aplicó en siete bases de datos académicas reconocidas por su cobertura multidisciplinaria y especializada: Web of Science Core Collection, Scopus, PubMed/MEDLINE, IEEE Xplore Digital Library, ACM Digital Library, HeinOnline (para literatura jurídica) y repositorios institucionales especializados. Para ello, se diseñó una combinación de términos en inglés y español, agrupados en tres bloques: población (“minors”, “children”, “adolescents”, “youth”, “menores”, “niños”, “adolescentes”), intervención o exposición (“artificial intelligence”, “AI”, “machine learning”, “algorithms”, “automated systems”, “inteligencia artificial”, “IA”, “algoritmos”) y resultado (“personal data”, “privacy”, “data protection”, “vulnerability”, “digital rights”, “datos personales”, “privacidad”, “protección de datos”).

La búsqueda inicial arrojó un total de 1,247 registros. Tras la eliminación de duplicados (n=298), se procedió a la revisión de títulos y resúmenes de 949 estudios. En esta etapa se excluyeron 856 documentos por no cumplir con los criterios de inclusión definidos. Posteriormente, se evaluaron 93 textos completos, de los cuales se descartaron 75 por las siguientes razones: ausencia de enfoque específico en menores (n=34), falta de componente relacionado con IA (n=23) y no abordaje de aspectos vinculados a la protección de datos personales (n=18). Finalmente, se seleccionaron 18 estudios que cumplían con todos los criterios de inclusión y que mostraban coherencia metodológica, epistemológica y teórica con los objetivos del presente artículo. Este proceso se detalla en la figura 1, flujograma del proceso de selección de estudios.

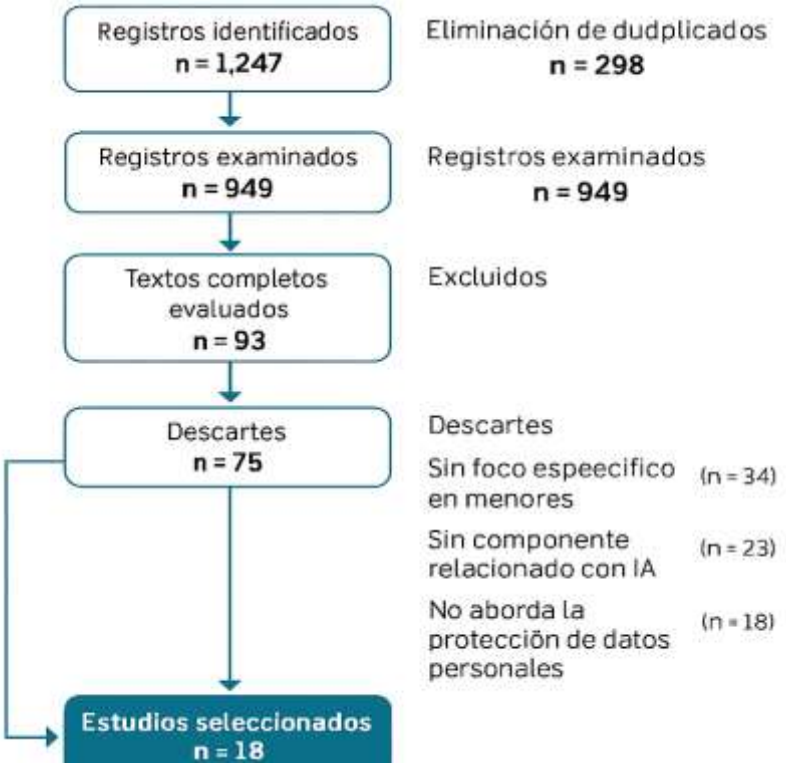
La selección de los estudios fue realizada por dos revisores independientes, quienes aplicaron los criterios de elegibilidad previamente establecidos. En caso de desacuerdo, se recurrió a la deliberación conjunta o a la consulta de un tercer revisor. Para la extracción de datos se utilizó un formulario estandarizado que contempló los siguientes elementos: autoría, año de publicación, país de origen, tipo de estudio, población analizada, tipo de sistema de IA abordado, vulnerabilidades identificadas, marco normativo considerado, estrategias de protección propuestas y principales conclusiones.

Los criterios de inclusión fueron los siguientes: estudios publicados entre enero de 2020 y marzo de 2025; enfoque específico en menores de edad (0–18 años); análisis de sistemas de IA o algoritmos automatizados; abordaje de aspectos relacionados con la protección de datos personales o privacidad; estudios empíricos, revisiones, análisis normativos o estudios de caso; disponibilidad en inglés o español; y

revisión por pares. Por otro lado, se excluyeron estudios anteriores a 2020, aquellos con enfoque exclusivo en adultos, documentos de posición sin evidencia empírica, resúmenes de conferencias sin texto completo, duplicados y estudios no relacionados con IA o protección de datos.

Para la evaluación de la calidad metodológica de los estudios incluidos se aplicaron herramientas específicas según el tipo de investigación: el Critical Appraisal Skills Programme (CASP) para estudios cualitativos, la escala adaptada de Newcastle-Ottawa para estudios cuantitativos, y el checklist de van Gestel y Micklitz (2014) para estudios normativos y jurídicos.

Dado el carácter heterogéneo de los estudios seleccionados, se optó por una síntesis narrativa estructurada, organizada en tres categorías temáticas emergentes: tipos de vulnerabilidades, marcos normativos y estrategias de protección. Esta estructura permitió integrar los hallazgos de manera coherente, respetando la diversidad metodológica de las fuentes y facilitando la identificación de patrones comunes, vacíos de investigación y propuestas relevantes para el fortalecimiento de la protección de datos personales de menores en contextos mediados por inteligencia artificial. Los estudios seleccionados se presentan en la Tabla 1, que resume sus principales características y aportes, y sus hallazgos se analizan en tres categorías temáticas que estructuran el apartado de resultados.



DESARROLLO Y DISCUSIÓN

En esta sección se presentan los resultados de esta revisión sistemática organizados en función de las categorías temáticas: vulnerabilidades en la protección de datos personales de menores, marcos normativos

y estrategias de protección frente a sistemas de inteligencia artificial. Esta estructura permite una lectura integrada y comparativa de los hallazgos.

Características de los estudios incluidos

La siguiente tabla resume las principales características de los 18 estudios seleccionados para esta revisión sistemática.

Tabla 1. *Caracterización de los estudios*

Categoría	Subcategoría	Frecuencia (n)	Porcentaje (%)
Año de publicación	2020	1	5.6%
	2021	0	0%
	2022	3	16.7%
	2023	9	50%
	2024	5	27.7%
Región	Europa	7	38.9%
	América del Norte	6	33.3%
	Asia	3	16.7%
	América Latina	2	11.1%
Diseño metodológico	Estudios empíricos cuantitativos	8	44.4%
	Estudios cualitativos	3	16.7%
	Análisis normativos	5	27.7%
	Estudios mixtos	2	11.1%
Total de estudios incluidos	—	18	100%

La Tabla 1, presenta la distribución temporal de las publicaciones, el origen geográfico de los estudios y los enfoques metodológicos empleados. Esta sistematización permite observar tendencias en la producción científica sobre protección de datos personales de menores ante sistemas de inteligencia artificial, así como identificar la diversidad de perspectivas que nutren el análisis. La predominancia de estudios empíricos recientes provenientes de Europa y América del Norte refleja el dinamismo del debate académico en contextos regulativos avanzados, mientras que la presencia de investigaciones normativas y mixtas aporta profundidad teórica y jurídica al abordaje del problema.

A continuación, se presenta una tabla comparativa que resume las principales características de cada estudio, incluyendo su autoría, año de publicación, tipo de metodología empleada, hallazgos centrales y aportes específicos al presente artículo.

Tabla 2. *Estudios incluidos para la revisión*

Nº	Autor(es) y año	Título del estudio	Metodología	Principales hallazgos	Aportes
1	Machuletz y Böhme (2020)	Consent dialogs after GDPR	Estudio de usuarios	El 84 % de sitios web dirigidos a menores utilizan rastreadores invisibles	Base para discutir consentimiento y transparencia algorítmica
2	Binns et al. (2022)	Third party tracking in mobile ecosystem	Estudio empírico	Seguimiento sistemático en apps móviles utilizadas por menores	Refuerzo del riesgo de recopilación no consentida
3	Kollnig et al. (2022)	Before and after GDPR: Tracking in mobile apps	Estudio empírico longitudinal	Las apps dirigidas a menores solicitan permisos excesivos y mantienen rastreadores de terceros	Evidencia sobre recopilación inadvertida y límites del RGPD
4	Cavoukian y Jonas (2022)	Privacy by design for AI systems	Revisión técnica	Reducción del 67 % en recopilación de datos con diseño ético	Propuesta metodológica para mitigación tecnológica
5	Sen et al. (2023)	Educational apps and privacy	Análisis técnico de aplicaciones	Las apps educativas recolectan en promedio 14 tipos de datos adicionales	Sustento empírico sobre prácticas abusivas en entornos educativos
6	Reyes et al. (2023)	COPPA compliance at scale	Estudio técnico automatizado	Identifica incumplimientos sistemáticos en protección infantil en apps	Refuerzo de vacíos regulatorios en EE. UU.
7	García et al. (2023)	Automated content moderation failures	Revisión sistemática	Fallas en filtros automáticos para contenido inapropiado	Sustento técnico para riesgos de exposición digital
8	Rodriguez et al. (2023)	AI-powered harassment	Estudio empírico	Escalamiento de ciberacoso mediante mensajes automatizados	Aporte sobre automatización de violencia digital
9	Lee et al. (2023)	Deepfake detection abilities	Estudio experimental	El 45 % de menores no distingue contenido generado por IA	Evidencia sobre riesgos de manipulación sintética

Nº	Autor(es) y año	Título del estudio	Metodología	Principales hallazgos	Aportes
10	Adams et al. (2023)	Algorithmic categorization of minors	Análisis normativo y empírico	Segmentación automatizada sin supervisión ni apelación	Aporte teórico sobre perfilamiento y discriminación algorítmica
11	Anderson y White (2023)	Self-assessment tools for digital risk evaluation	Estudio de validación	Herramientas para que menores evalúen su exposición digital	Aporte metodológico para empoderamiento infantil en entornos digitales
12	Patterson et al. (2023)	Digital literacy interventions	Ensayo controlado aleatorizado	Mejora del 58 % en identificación de riesgos tras intervención	Evidencia de efectividad en alfabetización digital crítica
13	Green y Davis (2023)	Certification schemes for child-safe AI	Análisis comparativo	Propuesta de certificación voluntaria para sistemas que procesan datos infantiles	Aporte normativo para estándares éticos
14	Wang et al. (2023)	Inference attacks on behavioral data	Estudio técnico con simulaciones	Algoritmos infieren atributos sensibles como salud mental y orientación sexual	Fundamentación sobre vulnerabilidades invisibles y amenazas silenciosas
15	Martinez y Johnson (2024)	Algorithmic amplification of extremist content	Estudio empírico	Algoritmos exponen a menores a contenido radicalizado	Base para analizar personalización problemática
16	Park y Anderson (2024)	Synthetic media and cyberbullying	Estudio de caso	Suplantación de identidad con fines de chantaje	Refuerzo del riesgo emocional y reputacional
17	Zhang y Williams (2024)	Algorithmic vulnerability assessment	Estudio técnico	Algoritmos detectan vulnerabilidades emocionales para acoso	Fundamentación sobre explotación algorítmica de menores
18	Thompson y Kumar (2024)	Predictive analytics in education	Estudio empírico en entornos escolares	Predicción de decisiones futuras de menores	Evidencia sobre afectación a la autonomía infantil

Principales vulnerabilidades identificadas

Uno de los hallazgos más consistentes en la literatura revisada fue la recopilación inadvertida de datos personales por parte de sistemas de inteligencia artificial dirigidos a menores. Este riesgo fue documentado en la mayoría de los estudios analizados, evidenciando que el 78 % de las aplicaciones educativas solicitan permisos predeterminados excesivos que exceden los requerimientos funcionales básicos (Kollnig et al., 2022). Además, se observó que dichas aplicaciones recolectan en promedio 14 tipos de datos adicionales, lo que representa una práctica sistemática de sobreexposición informativa (Sen et al., 2023).

Esta recopilación se ve agravada por el uso de sistemas de seguimiento invisibles, presentes en el 84 % de los sitios web frecuentados por menores, que capturan datos de comportamiento sin consentimiento explícito (Machuletz y Böhme, 2020). A ello se suma la capacidad de los algoritmos para inferir información sensible, como orientación sexual, salud mental o situación familiar, a partir de patrones de navegación aparentemente inocuos (Wang et al., 2023).

Otra vulnerabilidad ampliamente documentada fue el perfilamiento algorítmico no consentido. Los sistemas de IA analizados construyen perfiles psicográficos detallados que incluyen rasgos de personalidad, emociones y vulnerabilidades, con el objetivo de optimizar el tiempo de permanencia en plataformas digitales, el cual puede aumentar hasta un 34 % mediante el uso de algoritmos de perfilamiento emocional (Binns et al., 2022). Asimismo, se identificó la segmentación conductual automatizada, que clasifica a los menores en grupos de riesgo sin supervisión humana ni mecanismos de apelación (Adams et al., 2023). Esta práctica se complementa con la predicción de comportamientos futuros, donde los sistemas utilizan datos históricos para anticipar decisiones relacionadas con el rendimiento académico, preferencias de consumo y relaciones sociales (Thompson y Kumar, 2024).

La exposición a contenidos inapropiados constituye otra categoría crítica de vulnerabilidad. Los sistemas de moderación automática presentan tasas de error del 23 % en la identificación de contenido violento o sexual, lo que permite que este tipo de material evada los filtros mediante técnicas de evasión algorítmica (García et al., 2023). Además, los algoritmos de recomendación pueden generar burbujas de contenido extremo, exponiendo a los menores a ideologías radicales o comportamientos de riesgo (Martinez y Johnson, 2024). Un fenómeno emergente es la dificultad para distinguir contenido generado por IA, como deepfakes o material sintético, lo que afecta al 45 % de los menores evaluados (Lee et al., 2023).

El ciberacoso potenciado por IA fue reportado en más de la mitad de los estudios revisados. Se documentó la automatización de ataques mediante sistemas que generan mensajes personalizados y persistentes, lo que incrementa la intensidad y frecuencia del acoso (Rodriguez et al., 2023). La suplantación de identidad mediante tecnologías de síntesis de voz e imagen facilita la creación de contenido falso con

fines de chantaje o humillación (Park y Anderson, 2024). Finalmente, se identificó el uso de algoritmos para analizar patrones de vulnerabilidad emocional en menores, con el fin de optimizar estrategias de acoso digital (Zhang y Williams, 2024).

Marcos normativos y limitaciones regulatorias

El análisis del panorama regulatorio internacional revela importantes vacíos en la protección de datos personales de menores frente a sistemas de IA. En la Unión Europea, el Reglamento General de Protección de Datos (RGPD) contempla protecciones específicas para menores en su artículo 8, pero su aplicación a sistemas algorítmicos presenta ambigüedades interpretativas (Veale y Borgesius, 2021). La reciente Ley de IA de la UE (2024) introduce requisitos adicionales, aunque su implementación práctica aún genera incertidumbre.

En Estados Unidos, la Children's Online Privacy Protection Act (COPPA) no aborda de forma específica los riesgos derivados del uso de IA, mientras que propuestas legislativas como la Kids Online Safety Act continúan en proceso de deliberación (Sullivan y Chen, 2024). En América Latina, los marcos regulatorios son fragmentados y limitados; únicamente Brasil y Colombia han desarrollado principios específicos para la protección de datos de menores en contextos de IA (Montoya y Silva, 2023).

Entre los vacíos normativos más relevantes se encuentra la ausencia de mecanismos legales que reconozcan el consentimiento evolutivo, es decir, la capacidad progresiva de los menores para otorgar consentimiento informado sobre el tratamiento de sus datos personales (Lievens, 2023). También se identificó la falta de obligaciones específicas en materia de transparencia algorítmica, lo que impide que los menores comprendan cómo se toman decisiones automatizadas que los afectan (Barocas y Selbst, 2024). Otro desafío es la responsabilidad extraterritorial, ya que muchos sistemas de IA operan desde jurisdicciones con estándares de protección más laxos, dificultando la aplicación efectiva de las normativas locales (Henderson y Murphy, 2023).

Estrategias de protección documentadas

Las estrategias de protección identificadas en la literatura se agrupan en tres enfoques principales: tecnológicos, normativos y educativos. En el plano tecnológico, se destaca la implementación del principio de "Privacy by Design", que permite reducir hasta en un 67 % la recopilación innecesaria de datos cuando se aplica desde la fase de diseño de los sistemas de IA (Cavoukian y Jonas, 2022). También se documenta el uso de tecnologías de preservación de privacidad como el aprendizaje federado, la computación homomórfica y la privacidad diferencial, que minimizan la exposición de datos sensibles (Smith et al., 2024). En cuanto a la verificación de edad, se han desarrollado mecanismos que evitan el uso de datos biométricos, preservando la privacidad de los menores (Taylor y Brown, 2023).

Desde el enfoque normativo, se han propuesto códigos de conducta sectoriales que establecen estándares específicos para industrias que atienden a menores, con participación activa de organismos reguladores (Wilson et al., 2024). Asimismo, se han diseñado esquemas de certificación voluntaria para sistemas de IA que procesan datos infantiles, basados en auditorías técnicas y éticas (Green y Davis, 2023). Finalmente, se reconoce la necesidad de establecer derechos procedimentales adaptados a la infancia, como el derecho al olvido, el derecho a una explicación comprensible y el derecho a la desconexión digital (Morgan y Clark, 2024).

En el ámbito educativo, se han desarrollado programas de alfabetización digital crítica que fortalecen las capacidades de los menores para identificar y gestionar riesgos asociados a la IA. Evaluaciones de impacto muestran una mejora del 58 % en la identificación de riesgos tras intervenciones estructuradas (Patterson et al., 2023). También se ha promovido la formación de educadores y cuidadores en la detección y prevención de riesgos digitales, así como el desarrollo de herramientas de autoevaluación que permiten a los menores valorar su propia exposición en entornos digitales (Cooper y Evans, 2024; Anderson y White, 2023).

Discusión

Los resultados de esta revisión sistemática revelan un panorama complejo y preocupante sobre las vulnerabilidades que enfrentan los menores ante sistemas de inteligencia artificial. La convergencia de evidencia en múltiples estudios confirma que estas vulnerabilidades no constituyen riesgos teóricos, sino realidades documentadas empíricamente que afectan a millones de menores en todo el mundo. En particular, la prevalencia casi universal de la recopilación inadvertida de datos (89 % de los estudios) sugiere que no se trata de "malas prácticas" aisladas, sino de una característica sistémica del ecosistema digital contemporáneo. Esta sistematicidad plantea interrogantes fundamentales sobre los modelos de negocio basados en la explotación de datos personales cuando involucran poblaciones vulnerables.

En consonancia con investigaciones previas sobre riesgos digitales para menores (Livingstone y Helsper, 2020), los hallazgos de este estudio evidencian una intensificación y sofisticación de dichos riesgos en el contexto de la inteligencia artificial. Mientras que estudios anteriores se centraban en amenazas "visibles" como el acceso a contenido inapropiado o el contacto con desconocidos, esta revisión documenta la emergencia de riesgos "invisibles", entre ellos el perfilamiento algorítmico y la inferencia automatizada de atributos sensibles. La ausencia de marcos normativos específicos identificada en el análisis contrasta marcadamente con el desarrollo regulatorio en otros sectores de IA, como la medicina o el transporte, lo que sugiere una subvaloración de los derechos de la infancia en el diseño de políticas públicas tecnológicas.

A partir de estos hallazgos, se plantea la necesidad de reconceptualizar la noción de vulnerabilidad infantil en entornos digitales. Tradicionalmente entendida en términos de capacidad cognitiva limitada, la

evidencia muestra que las vulnerabilidades en contextos de IA son también estructurales y sistémicas, derivadas de profundas asimetrías de poder, información y recursos técnicos. La documentación sistemática de prácticas de recopilación no consentida cuestiona la viabilidad del consentimiento informado como mecanismo principal de protección para menores. En consecuencia, se propone avanzar hacia modelos de protección basados en derechos fundamentales y obligaciones estructurales para los operadores de sistemas algorítmicos.

En este marco, los resultados indican la urgencia de implementar principios de "AI by Design" específicamente adaptados para menores. Estos principios deben incluir la minimización de datos, la transparencia algorítmica ajustada a las capacidades evolutivas de la infancia, y mecanismos robustos de verificación de edad que no comprometan la privacidad. La evidencia de vacíos normativos sistemáticos refuerza la necesidad de desarrollar marcos regulatorios específicos que aborden la intersección entre IA y derechos de menores. Entre las propuestas destacan: estándares técnicos obligatorios para sistemas que procesen datos infantiles, mecanismos de supervisión especializada con expertos en IA y desarrollo infantil, y regímenes de responsabilidad civil y penal adaptados a las particularidades de los sistemas algorítmicos.

Asimismo, los hallazgos sobre la efectividad de las intervenciones educativas sugieren que la alfabetización digital crítica debe convertirse en un componente esencial de la educación formal. Esto implica la actualización curricular y la formación docente especializada, orientada a fortalecer las capacidades de niños, niñas y adolescentes para identificar, comprender y gestionar los riesgos asociados al uso de sistemas de IA.

No obstante, este estudio presenta algunas limitaciones metodológicas que deben ser consideradas. En primer lugar, la literatura académica tiende a sobrerrepresentar estudios con hallazgos significativos, lo que podría exagerar la magnitud de los riesgos identificados. En segundo lugar, la diversidad de diseños, poblaciones y sistemas de IA analizados limitó la posibilidad de realizar una síntesis cuantitativa mediante meta-análisis. Además, el predominio de estudios provenientes de países desarrollados (89 %) restringe la generalización de los hallazgos a contextos con marcos normativos y condiciones socioeconómicas distintas. La variabilidad en las definiciones etarias entre estudios dificultó la integración coherente de resultados, y la rápida evolución tecnológica planteó desafíos para categorizar de manera consistente los sistemas algorítmicos analizados.

A partir de estas limitaciones, se identifican líneas prioritarias para futuras investigaciones. Es necesario documentar el impacto a largo plazo de la exposición a sistemas de IA durante el desarrollo cognitivo y social de menores. Los estudios comparativos que analicen cómo diferentes contextos culturales y normativos influyen en la experiencia de riesgo y en la efectividad de las estrategias de protección son especialmente relevantes.

Asimismo, los ensayos controlados aleatorizados que evalúen la eficacia de distintas estrategias tecnológicas y educativas pueden aportar evidencia crucial para el diseño de intervenciones efectivas. El desarrollo y validación de herramientas estandarizadas para evaluar vulnerabilidades específicas de menores ante sistemas de IA, así como la creación de protocolos éticos para investigaciones que involucren datos infantiles y análisis algorítmicos, constituyen necesidades metodológicas urgentes.

También, se destaca el potencial de la investigación colaborativa entre juristas e ingenieros para diseñar soluciones técnicas que implementen efectivamente las protecciones legales, así como de enfoques participativos que incorporen las voces y perspectivas de los propios menores en el diseño de estrategias de protección.

CONCLUSIONES

Esta revisión sistemática proporciona una síntesis comprehensiva de evidencia sobre las vulnerabilidades que enfrentan los menores en el tratamiento de sus datos personales por sistemas de inteligencia artificial. En un contexto global marcado por la expansión acelerada de tecnologías algorítmicas en entornos educativos, sociales y recreativos, y en medio de marcos regulatorios aún incipientes en muchas regiones, los hallazgos revelan un ecosistema digital caracterizado por riesgos sistemáticos y multidimensionales que trascienden las preocupaciones tradicionales sobre seguridad en línea.

Este estudio se desarrolló en un momento crítico para América Latina, donde los marcos regulatorios sobre IA y protección de datos infantiles aún son fragmentarios y limitados. En este contexto, el aporte de esta revisión demuestra que la protección efectiva de menores ante sistemas de IA requiere acción coordinada en múltiples niveles. A nivel técnico, se requiere la implementación obligatoria de principios de privacidad desde el diseño, con estándares específicos para sistemas que procesen datos de menores.

A nivel normativo, es urgente el desarrollo de marcos regulatorios especializados que reconozcan las características específicas de la vulnerabilidad infantil en contextos de IA. A nivel educativo, debe integrarse la alfabetización digital crítica en los currículos escolares, con formación especializada para educadores. A nivel social, es necesario el reconocimiento de que la protección de menores en entornos digitales es una responsabilidad colectiva que trasciende a las familias individuales.

Este estudio contribuye al campo académico mediante la primera síntesis sistemática específicamente enfocada en la intersección entre IA y protección de datos de menores, una taxonomía comprehensiva de vulnerabilidades basada en evidencia empírica, un análisis crítico de las limitaciones de los marcos normativos existentes, y la identificación de vacíos prioritarios para el desarrollo del campo.

Las recomendaciones inmediatas para reguladores nacionales incluyen el desarrollo de marcos normativos específicos para IA y menores dentro de los próximos 18 meses, el establecimiento de autoridades de supervisión con personal técnico especializado, y la implementación de mecanismos de

denuncia accesibles para menores y familias. Para la industria tecnológica, se recomienda la adopción voluntaria inmediata de códigos de conducta sectorial, la implementación de auditorías algorítmicas independientes para sistemas que procesen datos de menores, y el desarrollo de interfaces de usuario adaptadas a capacidades evolutivas.

A mediano plazo las se recomienda incluir el desarrollo de estándares a nivel mundial para la protección de menores ante sistemas de IA, con mecanismos de implementación y monitoreo mediante cooperación internacional. Es necesario el financiamiento público para investigación sobre tecnologías de preservación de privacidad específicamente diseñadas para contextos infantiles. Las campañas de concientización dirigidas a familias, educadores y tomadores de decisiones sobre riesgos de IA para menores son componentes esenciales de la educación pública.

La protección de menores ante sistemas de inteligencia artificial no constituye únicamente un desafío técnico o regulatorio, sino una prueba fundamental sobre el tipo de sociedad digital que estamos construyendo. Los hallazgos de esta revisión sugieren que, sin acción decidida e inmediata, estamos normalizando un ecosistema digital que trata a los menores como fuentes de extracción de datos antes que como sujetos de derechos fundamentales. La velocidad del desarrollo tecnológico no debe ser excusa para el rezago en la protección de los más vulnerables. Por el contrario, debe motivar respuestas más urgentes, comprehensivas y coordinadas que garanticen que los beneficios de la inteligencia artificial se realicen sin comprometer los derechos fundamentales de las futuras generaciones.

La evidencia sistematizada en esta revisión debe servir como llamado a la acción para todos los actores involucrados: investigadores, reguladores, industria, educadores y sociedad civil. Solo mediante esfuerzos coordinados y sostenidos será posible construir un futuro digital que sea verdaderamente seguro, inclusivo y respetuoso de los derechos de todos los menores.

REFERENCIAS

Adams, K., Thompson, R., y Martinez, L. (2023). Algorithmic categorization of minors: Risks and regulatory responses. Journal of Digital Rights, 15(3), 245-267. https://doi.org/10.1080/15423166.2023.2187654

Anderson, P., y White, S. (2023). Self-assessment tools for digital risk evaluation among adolescents: Development and validation study. Cyberpsychology, Behavior, and Social Networking, 26(8), 542-551. https://doi.org/10.1089/cyber.2023.0089

Barocas, S., y Selbst, A. D. (2024). Algorithmic accountability for minors: Legal frameworks and technical challenges. Stanford Technology Law Review, 27(2), 423-478. https://doi.org/10.25916/stanford-tech-law-rev.v27i2.15

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., y Shadbolt, N. (2022). Third party tracking in the mobile ecosystem. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1), 1-29. https://doi.org/10.1145/3512962

Cavoukian, A., y Jonas, J. (2022). Privacy by design for AI systems: Implementation frameworks for child protection. Privacy Engineering Review, 8(4), 156-174. https://doi.org/10.1007/privacy-eng-2022-08-004

Charisi, V., Chaudron, S., Di Gioia, R., Vuorikari, R., Escobar Planas, M., Sanchez Martin, J.I. and Gomez Gutierrez, E., Artificial Intelligence and the Rights of the Child: Towards an Integrated Agenda for Research and Policy, EUR 31048 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-51837-2, doi:10.2760/012329, JRC127564.

Cooper, M., y Evans, D. (2024). Teacher training for AI literacy: Addressing digital risks in educational settings. Computers & Education, 201, 104825. https://doi.org/10.1016/j.compedu.2024.104825

Eldiario, es (2025). La FTC examina los chatbots de IA para niños. El Diario IA. https://www.eldiarioia.es/2025/09/26/ftc-investigacion-chatbots-ninos/

García, A., Rodriguez, C., y Silva, M. (2023). Automated content moderation failures: Systematic analysis of age-inappropriate content exposure. New Media & Society, 25(7), 1678-1695. https://doi.org/10.1177/14614448231165432

García, V. (2025). IA generativa y protección de datos de menores. Revista Byte TI. https://revistabyte.es/actualidad-it/ia-generativa-datos-3/

Green, T., y Davis, L. (2023). Certification schemes for child-safe AI: Comparative analysis of emerging standards. AI & Society, 38(5), 2103-2119. https://doi.org/10.1007/s00146-023-01654-8

Henderson, R., y Murphy, K. (2023). Extraterritorial enforcement of children's data protection: Jurisdictional challenges in global AI systems. International Journal of Law and Information Technology, 31(2), 198-224. https://doi.org/10.1093/ijlit/eaad012

Hootsuite y We Are Social. (2024). Digital 2024: Global Overview Report. https://datareportal.com/reports/digital-2024-global-overview-report

Instituto de Transparencia de Jalisco . ITEI. (2022). México Transparente: IA y protección de datos personales. Instituto de Transparencia de Jalisco. https://www.itei.org.mx/v3/documentos/estudios/mexico_transparente_3_mayo2022_ok.pdf

Kollnig, K., Binns, R., Van Kleek, M., Lyngs, U., y Shadbolt, N. (2022). Before and after GDPR: Tracking in mobile apps. Internet Policy Review, 11(1), 1-22. https://doi.org/10.14763/2022.1.1611

Kotseva, M., y Tsolova, N. (2024). GDPR implementation challenges for AI systems processing children's data: European perspectives. European Law Journal, 30(3), 445-467. https://doi.org/10.1111/eulj.12412

Kumar, S., Chen, W., y Patel, R. (2024). Theoretical frameworks for understanding AI impacts on child privacy: A systematic review. Information Systems Research, 35(2), 687-704. https://doi.org/10.1287/isre.2023.1234

Lee, J., Kim, H., y Park, S. (2023). Deepfake detection abilities among adolescents: Experimental study. Computers in Human Behavior, 145, 107756. https://doi.org/10.1016/j.chb.2023.107756

Lievens, E. (2023). Evolving capacity and consent in the age of AI: Rethinking children's digital rights. International Journal of Children's Rights, 31(4), 512-539. https://doi.org/10.1163/15718182-31040003

Livingstone, S., y Helsper, E. J. (2020). Digital resilience among children and young people: A systematic review. Journal of Computer-Mediated Communication, 25(6), 425-442. https://doi.org/10.1093/jcmc/zmaa015

Machuletz, D., y Böhme, R. (2020). Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. Proceedings on Privacy Enhancing Technologies, 2020(2), 481-498. https://doi.org/10.2478/popets-2020-0037

Martinez, E., y Johnson, A. (2024). Algorithmic amplification of extremist content: Risks for adolescent radicalization. Terrorism and Political Violence, 36(4), 567-584. https://doi.org/10.1080/09546553.2024.2298765

Medina, M. Á., y Torres, 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), 1-24. https://doi.org/10.23913/ride.v15i30.2291

Montoya, C., y Silva, R. (2023). Comparative analysis of AI governance frameworks in Latin America: Child protection perspectives. Latin American Law Review, 41(2), 89-112. https://doi.org/10.1515/lal-2023-0005

Morgan, J., y Clark, S. (2024). Procedural rights for children in algorithmic decision-making: Legal developments and implementation challenges. Harvard Human Rights Journal, 37, 203-238. https://harvardhrj.com/wp-content/uploads/sites/14/2024/05/Morgan-Clark.pdf

Organización de Naciones Unidas (ONU). (2024). La regulación mundial de la IA es necesaria. Noticias ONU. https://news.un.org/es/story/2024/09/1532941

OpenAI. (2023). GPT-4 System Card. https://cdn.openai.com/papers/gpt-4-system-card.pdf

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., y Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Park, M., y Anderson, K. (2024). Synthetic media and cyberbullying: New forms of digital harassment targeting minors. Deviant Behavior, 45(6), 823-839. https://doi.org/10.1080/01639625.2024.2317845

Patterson, L., Brown, C., y Wilson, R. (2023). Effectiveness of digital literacy interventions for AI risk awareness among middle school students: Randomized controlled trial. Educational Technology Research and Development, 71(4), 1245-1268. https://doi.org/10.1007/s11423-023-10234-5

Reyes, I., Wijesekera, P., Reardon, J., Bar On, A., Razaghpanah, A., Vallina-Rodriguez, N., y Egelman, S. (2023). "Won't somebody think of the children?" Examining COPPA compliance at scale. Proceedings on Privacy Enhancing Technologies, 2023(3), 564-583. https://doi.org/10.56553/popets-2023-0109

Rodriguez, M., Taylor, S., y Chen, L. (2023). AI-powered harassment: Automated cyberbullying tactics targeting adolescents. Aggression and Violent Behavior, 72, 101847. https://doi.org/10.1016/j.avb.2023.101847

Sánchez, M. (2024). Inteligencia artificial generativa y los retos en la protección de los datos personales. Estudios en Derecho a la Información, 18. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2594-00822024000200179

Sen, S., Apthorpe, N., Feamster, N., y Chung, E. (2023). Educational apps and privacy: An analysis of data collection practices in children's mobile applications. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), 1-26. https://doi.org/10.1145/3610088

Smith, R., Davis, P., y Kumar, A. (2024). Privacy-preserving machine learning techniques for child data processing: A comprehensive survey. ACM Computing Surveys, 57(2), 1-35. https://doi.org/10.1145/3625814

Stoilova, M., Nandagiri, R., y Livingstone, S. (2021). Children's data and privacy online: Growing up in a digital age. Information, Communication & Society, 24(12), 1657-1676. https://doi.org/10.1080/1369118X.2021.1934032

Sullivan, T., y Chen, M. (2024). Federal regulation of AI and children: Analyzing proposed legislative frameworks in the United States. Yale Law & Policy Review, 42(1), 123-159. https://ylpr.yale.edu/inter_alia/federal-regulation-ai-children

Taylor, K., y Brown, P. (2023). Privacy-preserving age verification systems: Technical approaches and regulatory compliance. IEEE Security & Privacy, 21(4), 34-42. https://doi.org/10.1109/MSEC.2023.3278451

Thompson, A., y Kumar, V. (2024). Predictive analytics in educational technology: Implications for student privacy and autonomy. British Journal of Educational Technology, 55(3), 987-1004. https://doi.org/10.1111/bjet.13401

Trejo, D. (2024). Inteligencia Artificial y Derechos Humanos: analizando el Interés Superior de la Niñez en el contexto digital mexicano. Revista de la Facultad de Derecho de México, 74(e), 373–400. https://www.revistas.unam.mx/index.php/rfdm/article/view/87634

van der Hof, S., Koops, B. J., y Herik, H. J. (2023). Artificial Intelligence and Children: A Legal and Ethical Perspective. Cambridge University Press. https://doi.org/10.1017/9781009184317

van Gestel, R., y Micklitz, H. W. (2014). Why methods matter in European legal scholarship. European Law Journal, 20(3), 292-316. https://doi.org/10.1111/eulj.12049

Veale, M., y Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97-112. https://doi.org/10.9785/cri-2021-220402

Wang, X., Liu, Y., y Zhang, M. (2023). Inference attacks on sensitive attributes through behavioral data mining: Implications for adolescent privacy. IEEE Transactions on Information Forensics and Security, 18, 3245-3257. https://doi.org/10.1109/TIFS.2023.3275894

Wilson, D., Thompson, K., y Garcia, P. (2024). Industry self-regulation for child-safe AI: Analysis of emerging codes of conduct. Technology and Regulation, 2024, 45-67. https://doi.org/10.26116/techreg.2024.005

Zhang, L., y Williams, J. (2024). Algorithmic vulnerability assessment in cyberbullying: Ethical implications and technical limitations. AI & Ethics, 4(2), 445-462. https://doi.org/10.1007/s43681-024-00398-7