Revista Iberoamericana de la Propiedad Intelectual
N° 22 (2025): 123-154

Sesgos en la inteligencia artificial en el sector salud: una revisión sistemática en el contexto iberoamericano

* * * *

Jéssica Bárcenas Vidal[1]
Universidad Católica de Temuco
jbarcenas@uct.cl
https://orcid.org/0000-0001-5759-2682

Isnel Martínez Montenegro
Universidad Católica de Temuco
imartinez@uct.cl
https://orcid.org/0000-0003-0322-1071

Carlos Fuertes Iglesias
Universidad de Zaragoza
cfuertes@unizar.es
https://orcid.org/0000-0002-0791-2390

Recibido: 15 de mayo de 2025
Aceptado: 24 de junio de 2025

Resumen
La inteligencia artificial (IA) ha emergido como una herramienta transformadora en el ámbito de la salud, ofreciendo la promesa de mejorar diagnósticos, optimizar tratamientos y fomentar una atención médica más eficiente. En el contexto de la Agenda 2030 para el Desarrollo Sostenible, la adopción de tecnologías digitales es fundamental

1 Este artículo es parte de la investigación de sus estudios de doctorado en Derechos Humanos y Libertades Fundamentales de la Universidad de Zaragoza.

para alcanzar la cobertura sanitaria universal y proteger a las comunidades durante emergencias sanitarias. Sin embargo, la expansión de la IA en salud ha suscitado preocupaciones sobre la posibilidad de que perpetúe o amplifique sesgos preexistentes. La investigación en este campo ha predominado en países anglosajones, pero deja un vacío sobre el estado de la IA en salud en Iberoamérica. De esta forma, este artículo es una revisión sistemática que tiene como objetivo analizar la literatura científica sobre el uso y los sesgos de la IA en el sector salud en países iberoamericanos y latinoamericanos, con el fin de identificar las principales áreas de aplicación, los desafíos y las brechas de investigación en comparación con los estudios realizados en países anglosajones. Asimismo, se examinan las implicaciones jurídicas y éticas de estos sesgos, considerando la necesidad de desarrollar marcos regulatorios que protejan los derechos de los pacientes y promuevan la equidad en la atención médica. Finalmente, se proponen recomendaciones para futuras investigaciones y el desarrollo de políticas que aborden la representación de datos y la transparencia en el uso de IA en salud, con el fin de mitigar los sesgos y garantizar una atención médica justa y equitativa.

Palabras clave: inteligencia artificial, atención médica, sesgos algorítmicos, inteligencia artificial en salud, sesgos en salud.

Biases in Artificial Intelligence in the Health Sector: A Systematic Review in the Ibero-American Context

Abstract
Artificial Intelligence (AI) has emerged as a transformative tool in the healthcare field, offering the promise of improving diagnostics, optimizing treatments, and promoting more efficient medical care. In the context of the 2030 Agenda for Sustainable Development, the adoption of digital technologies is essential to achieving universal health coverage and protecting communities during health emergencies. However, the expansion of AI in health has raised concerns about the possibility of perpetuating or amplifying pre-existing biases. Research in this field has been dominated by Anglo-Saxon countries, leaving a gap regarding the status of AI in health in Ibero-America. Thus, this article is a systematic review aimed at analyzing the scientific literature on the use and biases of AI in the health sector in Ibero-American and Latin American countries, in order to identify the main areas of application, challenges, and research gaps compared to studies conducted in Anglo-Saxon countries. It also examines the legal and ethical implications of these biases, considering the need to develop regulatory frameworks that protect patients' rights and promote equity in healthcare. Finally, recommendations are proposed for future research and policy development to address data representation and transparency in the use of AI in health, with the goal of mitigating biases and ensuring fair and equitable medical care.

Key words: artificial intelligence, medical care, algorithmic bias, artificial intelligence in health, health biases.

Vieses na inteligência artificial no setor de saúde: uma revisão sistemática no contexto ibero-americano

Resumo
A inteligência artificial (IA) surgiu como uma ferramenta transformadora no campo da saúde, oferecendo a promessa de melhorar diagnósticos, otimizar tratamentos e promover um atendimento médico mais eficiente. No contexto da Agenda 2030 para o Desenvolvimento Sustentável, a adoção de tecnologias digitais é fundamental para alcançar a cobertura universal de saúde e proteger as comunidades durante emergências sanitárias. No entanto, a expansão da IA na saúde tem gerado preocupações sobre a possibilidade de perpetuar ou amplificar vieses preexistentes. A pesquisa nesta área tem sido predominante em países anglo-saxônicos, deixando uma lacuna sobre o estado da IA em saúde na Ibero-América. Assim, este artigo é uma revisão sistemática que tem como objetivo analisar a literatura científica sobre o uso e os vieses da IA no setor da saúde em países ibero-americanos e latino-americanos, com o intuito de identificar as principais áreas de aplicação, desafios e lacunas de pesquisa em comparação com os estudos realizados em países anglo-saxônicos. Também são examinadas as implicações jurídicas e éticas desses vieses, considerando a necessidade de desenvolver marcos regulatórios que protejam os direitos dos pacientes e promovam a equidade no atendimento médico. Por fim, são propostas recomendações para futuras pesquisas e para o desenvolvimento de políticas que abordem a representação de dados e a transparência no uso da IA na saúde, visando mitigar os vieses e garantir um atendimento médico justo e equitativo.

Palavras-chave: inteligência artificial, atendimento médico, viés algorítmico, inteligência artificial na saúde, vieses na saúde.

1. Introducción

La agenda 2030 para el Desarrollo Sostenible establece que la expansión de las tecnologías de la información, junto con la interconexión mundial, ofrece la posibilidad de acelerar el progreso humano en diversas áreas. En el ámbito de la salud, las tecnologías digitales se han consolidado como esenciales para alcanzar la cobertura sanitaria universal y proteger a las personas en casos de emergencias sanitarias, así como para promover la salud y el bienestar (Estrategia OMS). Al respecto, la Organización Mundial de la Salud (OMS, 2021a) ha fomentado activamente la digitalización y el uso de las tecnologías, proporcionando asesoría técnica y normativa.

Dicho avance tecnológico sin duda fue potenciado activamente con la irrupción de la pandemia por COVID-19. En dicho contexto, la IA se usó masivamente para el análisis de grandes volúmenes de datos con el fin de predecir futuros brotes, modelar la propagación de la enfermedad y, en ciertos casos, para diagnosticarla (Cosgrove et al., 2020; Davis, 2020; Sekalala et al., 2020; Sun et al., 2020). Más aún, una encuesta de Market.us ha dicho que la IA generativa en la industria de la salud alcanzará alrededor de $17.000 millones para 2032, impulsada principalmente por la automatización en las operaciones de salud, imágenes médicas y diagnósticos, así como en la investigación y desarrollo de fármacos (Sai et al., 2024).

En paralelo, el uso de algoritmos de inteligencia artificial (IA) en el ámbito de la salud ha generado grandes expectativas sobre su potencial para mejorar la precisión diagnóstica, optimizar tratamientos y promover una atención médica más eficiente (Byrne, 2021). En concreto, las plataformas de IA en atención médica se están desarrollando para ser utilizadas en áreas como diagnósticos, monitoreo de pacientes y apoyo a la toma de decisiones tanto clínicas como de políticas públicas (Lysaght et al., 2019). Por ejemplo, se ha utilizado IA en el ámbito de la medicina de precisión, un enfoque que personaliza la atención médica según las características individuales de cada paciente, como su genética y factores ambientales. En este ámbito, la IA ha facilitado el análisis de grandes volúmenes de datos clínicos y genómicos para mejorar diagnósticos y tratamientos. Dicha convergencia les permite a los médicos ofrecer decisiones más informadas al combinar datos estructurados y no estructurados, optimizando así la atención (Johnson et al., 2021).

Otro campo en el que la IA ha tenido avances significativos ha sido en la radiología diagnóstica, enfocándose en su capacidad para mejorar la precisión y eficiencia en la interpretación de imágenes médicas, como radiografías, resonancias magnéticas y tomografías computarizadas. Entre sus principales beneficios, se encuentran la detención de patrones complejos y anomalías en las imágenes que, a menudo, son difíciles de identificar para los radiólogos humanos.

Como consecuencia de la propagación de la IA en Salud, han surgido preocupaciones sobre los efectos nocivos que ellas pueden tener en las personas. Particularmente, este texto abordará la inquietud so-

bre la posibilidad de que estas tecnologías perpetúen o amplifiquen sesgos existentes, afectando de manera desproporcionada a algunos grupos (Chen et al., 2023), como las minorías raciales, de género y socioeconómicas, entre otras. Por ejemplo, en 2019 se evidenció que un algoritmo priorizó a pacientes blancos sobre negros dependiendo de los costos de atención proyectados, lo que reflejó desigualdades en la atención. En otros casos, los algoritmos clínicos ajustados por raza excluyeron a minorías de procedimientos críticos, lo que trajo consigo el aumento de la mortalidad en pacientes negros (Byrne, 2021).

Sin lugar a dudas, este análisis requiere abordar el problema desde un enfoque jurídico, pues adquiere una relevancia crucial para abordar tanto las oportunidades como los riesgos asociados con el uso de la IA en el ámbito médico y sanitario (Ramón Fernández, 2021). A medida que la IA se integra en los sistemas de salud, se plantea la necesidad urgente de desarrollar marcos legales robustos que regulen su aplicación, de manera que se garanticen los derechos fundamentales de los pacientes, en particular su privacidad y la confidencialidad de la información médica. Al respecto, un informe de la Organización Mundial de la Salud (2021b) destaca la importancia de integrar principios éticos y derechos humanos en el desarrollo de tecnologías de IA para la salud, abogando por un enfoque en el que los derechos de los pacientes y comunidades sean priorizados para evitar la exacerbación de desigualdades en el acceso a la atención médica y la atención adecuada de los datos.

La regulación sobre estas relaciones médico-paciente debe abordar la responsabilidad en el uso de algoritmos, especialmente en lo que respecta a la toma de decisiones automatizadas que puedan afectar directamente la salud de las personas. En muchos países latinoamericanos, la regulación del uso de la IA en el sector salud aún está en sus etapas iniciales, lo que genera una brecha significativa en la protección jurídica de los pacientes frente a posibles discriminaciones o vulneraciones de derechos a causa de sesgos en los algoritmos utilizados en el servicio médico (Breceda Pérez y Castillo Lara, 2023).

El impacto de la propiedad intelectual en la expansión de la inteligencia artificial en salud también es un aspecto clave que no puede ser ignorado. Si bien las patentes y otros derechos de propiedad intelectual fomentan la innovación tecnológica, igualmente pueden ge-

nerar barreras significativas para el acceso equitativo a las tecnologías de la IA en sectores cruciales como la salud (Cruz Rivero, 2024). En particular, las grandes corporaciones tecnológicas que desarrollan algoritmos y sistemas de IA en salud pueden ejercer un control exclusivo sobre las herramientas que, en muchos casos, son esenciales para mejorar la atención médica (Viniegra-Velázquez, 2024). El control sobre estas innovaciones tecnológicas puede restringir el acceso a los sistemas de IA en países en vías de desarrollo, particularmente en Latinoamérica, donde las infraestructuras sanitarias no siempre tienen los recursos necesarios para implementar estas tecnologías de manera universal (García Uribe, 2022).

Incluso, la introducción de la IA en el sector salud exige una revisión crítica de las normativas que rigen el consentimiento informado y la autonomía del paciente. A medida que los algoritmos de IA desempeñan un papel cada vez más importante en los diagnósticos y la toma de decisiones médicas, surge la cuestión de cómo los pacientes pueden comprender y dar su consentimiento respecto al uso de estos sistemas (Azuaje Pirela, 2023).

Es relevante destacar que la problemática que aborda el presente artículo se ha desarrollado principalmente en países anglosajones (y otros), donde los sistemas de salud y las poblaciones estudiadas difieren significativamente de otros contextos. Por ejemplo, en los países latinoamericanos, los marcos legales sobre la protección de datos personales y el consentimiento en el ámbito de la salud aún no están completamente adaptados a los avances tecnológicos, lo que genera incertidumbre sobre la transparencia y la legitimidad del uso y control de los datos. Por ello, la necesidad de establecer fundamentos jurídicos que orienten un marco jurídico adaptado a estos desafíos es más urgente que nunca, a fin de garantizar que los avances tecnológicos en el sector de la salud no se traduzcan en nuevas formas de desigualdad o vulneración de derechos en los países latinoamericanos (Azuaje Pirela, 2023). Por lo tanto, el primer objetivo de este trabajo es analizar la literatura científica sobre el uso y los sesgos de la IA en el sector salud en países latinoamericanos, con el fin de identificar las principales áreas de aplicación, los desafíos y las brechas de investigación en comparación con los estudios realizados en países anglosajones. De igual modo, al ser la literatura científica en

la materia todavía incipiente en Latinoamérica, el estudio ampliará su análisis a los países iberoamericanos, término que permitiría un alcance más amplio de los estudios analizados.

A partir de los resultados obtenidos en la revisión sistemática, este estudio abordará las implicaciones jurídicas y éticas de los sesgos algorítmicos en el ámbito de la salud en los países bajo análisis y se proporcionarán conclusiones generales que servirán como base para futuras investigaciones sobre el tema. Finalmente, la discusión proporcionará un breve examen comparativo con estudios previos provenientes de países anglosajones que ya han identificado las implicaciones legales del sesgo en los sistemas de IA en salud.

2. Métodos

En este apartado de análisis se adoptó el enfoque de revisión sistemática PRISMA 2020, siguiendo sus elementos recomendados para la presentación de informes. Dicho recurso está diseñado para mejorar la calidad de los informes de revisiones sistemáticas y fomentar decisiones basadas en la evidencia (Page et al., 2021).

En cuanto a la búsqueda de artículos, se realizó en cuatro bases de datos digitales reconocidas: Scopus, Web of Science (WoS), SciELO y PubMed. Estas bases de datos fueron seleccionadas por su amplia cobertura de revistas médicas, científicas y de ciencias sociales en campos multidisciplinarios. En primer lugar, Scopus se destaca por sus recursos confiables y revisados por pares en diversos campos, incluyendo ciencias sociales, jurídicas, medicina y salud. En segundo lugar, Web of Science ofrece un recurso interdisciplinario que abarca artículos de investigación en ciencia, tecnología, salud y ciencias sociales. En tercer lugar, SciELO es un repositorio multidisciplinario; fue incluido por su enfoque en países iberoamericanos, especialmente en América Latina y el Caribe, mejorando la visibilidad y el acceso a la literatura científica en países en desarrollo. En cuarto lugar, PubMed es una base de datos de acceso libre especializada en ciencias de la salud; contiene más de 19 millones de referencias bibliográficas y es considerada una fuente clave en el campo médico debido a su extensa cobertura temática y actualización constante. En definitiva, la selección de estas bases de datos asegura una cobertura integral de

la investigación sobre sesgos algorítmicos en inteligencia artificial y salud en Iberoamérica.

Agregado a lo anterior, se establecieron criterios de inclusión y exclusión específicos para garantizar la relevancia y calidad de los estudios analizados en la revisión, por ende, se seleccionaron aquellos que abordan el uso de algoritmos de IA en el ámbito de la salud. En consecuencia, la búsqueda se realizó en las bases de datos Scopus, Web of Science, PubMed y SciELO, donde se utilizaron combinaciones de las siguientes palabras clave: "Artificial Intelligence" AND "Healthcare" AND "Bias" o "Artificial Intelligence" AND "Healthcare" AND "Algorithmic Bias".

Por otro lado, los artículos seleccionados debían estar indexados en dichas bases de datos y no se estableció ningún límite temporal en particular para su inclusión.

Como resultado, el uso de estas palabras clave proporcionó una recopilación inicial de 2.159 documentos. Posteriormente, se aplicaron criterios adicionales para refinar los resultados, enfocándose en estudios realizados en países iberoamericanos o aquellos que incluyeran datos relevantes sobre esta región, escritos en español o inglés. Asimismo, se excluyeron estudios que no estuvieran directamente relacionados con el uso de la IA en la salud o que no evaluaran sus sesgos asociados. Además, se descartaron documentos provenientes de fuentes no confiables, como aquellos no indexados o estudios teóricos sin muestra, así como otras revisiones de literatura. También se excluyeron investigaciones centradas únicamente en países fuera de Iberoamérica y artículos en idiomas distintos al español o inglés. Lo anterior redujo la búsqueda a 72 artículos.

Es importante señalar que, aunque se realizó una búsqueda utilizando una combinación de palabras clave en español en las mismas bases de datos, esta no arrojó resultados. Tras examinar los títulos y resúmenes de las publicaciones para verificar el uso eficaz de los términos de búsqueda, se identificaron 50 publicaciones pertinentes al propósito de este trabajo. Posteriormente, se evaluó el contenido de estos artículos, identificándose 14 estudios que cumplían plenamente con los criterios de pertinencia y relevancia para el objetivo de la investigación. Estos artículos, considerados "adecuados" por su contribución significativa al tema en cuestión, fueron incluidos en el

análisis final y se agruparon utilizando el procesador de datos Excel, donde se crearon tablas para facilitar el análisis presentado en este trabajo (ver Figura 1).

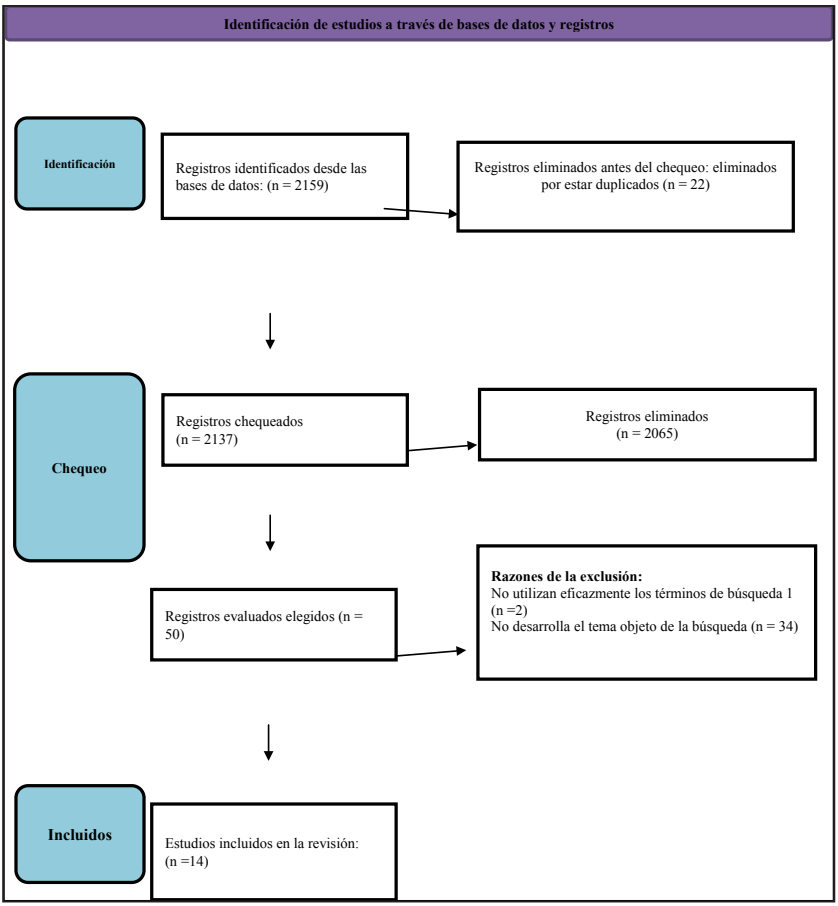

Figura 1: diagrama de flujo PRISMA para nuevas revisiones sistemáticas que incluyen búsquedas solo en bases de datos y registros. Fuente: guía de plantillas y uso del diagrama de flujo PRISMA 2020 para revisiones sistemáticas.

Finalmente, para abordar las implicaciones jurídicas de los sesgos algorítmicos en el área de la salud en los países bajo estudio, se utilizó una metodología de investigación cualitativa, combinada con un enfoque descriptivo y analítico. De acuerdo con Martínez Montenegro (2023), los métodos cualitativos permiten explorar las

percepciones, experiencias y realidades contextuales de los actores involucrados, lo que resulta especialmente relevante cuando se abordan temas complejos como la implementación de la IA en sectores sensibles como la salud. Esta metodología será útil para comprender no solo los avances tecnológicos, sino también las preocupaciones éticas y sociales relacionadas con los sesgos de los algoritmos.

3. Resultados

3.1 Innovaciones específicas y generativas de la IA en salud

En este apartado se presentan las aplicaciones de la IA en el ámbito salud que fueron abordadas en los estudios seleccionados. En ese sentido, se observa que la IA —y su implementación en salud— ha sido tratada de manera amplia en la literatura, sin embargo, algunos análisis se enfocaron en campos específicos donde se ha aplicado esta tecnología.

Al respecto, se mencionaron aplicaciones como la endoscopia asistida por IA, la cápsula endoscópica, la colonoscopia virtual y el análisis de muestras de heces, realzando cómo la integración de estas tecnologías ha transformando el diagnóstico y tratamiento de enfermedades en ese campo (Mascarenhas et al., 2023, 2024). En áreas como el diagnóstico radiológico, se mencionó que algunas aplicaciones de IA han mejorando el análisis de imágenes y la detección temprana de patologías mamarias (Rueda et al., 2024).

En campos como la nefrología, se destacaron aplicaciones de IA que utilizaron algoritmos para lograr identificar signos tempranos de enfermedades renales crónicas. Por ejemplo, los investigadores imitaron la capacidad de los nefropatólogos para extraer información diagnóstica, pronóstica y terapéutica de biopsias renales utilizando reconocimiento de imágenes (Fayos de Arizón et al., 2023).

En el ámbito de la atención en salud mental, Wilson et al. (2023) expusieron acerca de sistemas de IA utilizados para reducir los tiempos de espera en los departamentos de emergencia de los hospitales o como soporte a la toma de decisiones. Del mismo modo, ejemplificó la existencia de innovaciones digitales para abordar problemas como la soledad y la ansiedad, como robots de compañía para niños o personas mayores.

En otras áreas, se explicó que la IA está comenzando a influir en la asignación de recursos sanitarios escasos, en particular la asignación de hígados para trasplante. Tradicionalmente, dichos órganos se han asignado por criterios de urgencia, priorizando a aquellos en mayor riesgo de mortalidad a corto plazo. Sin embargo, la introducción de algoritmos de IA más sofisticados, como el sistema de beneficio de trasplante, logran mejorar, entre otras cosas, la precisión en la adaptación donante-receptor (Rueda et al., 2024).

Por otra parte, se evidencia el uso de la inteligencia artificial generativa (GAI, por sus siglas en inglés), que abarca sistemas capaces de crear contenido —como texto e imágenes— a partir de instrucciones humanas. Ejemplos notables incluyen modelos como DALL-E, que facilitan la detección y diagnóstico de enfermedades al analizar imágenes de pacientes, como rayos X y resonancias magnéticas, identificando patrones y anomalías que podrían pasar desapercibidos para los profesionales de la salud. Asimismo, al utilizar técnicas de generación de datos, pueden crearse imágenes que mejoran la precisión del diagnóstico y optimizan la intervención médica temprana. Entre otras cosas, también se evidencia que la GAI puede aplicarse en la educación médica, brindando a los profesionales escenarios virtuales para practicar y refinar habilidades clínicas. También existen modelos de lenguaje grandes (LLMs), que generan respuestas coherentes y contextualmente relevantes, y los modelos generativos adversariales (GANs), que producen imágenes y datos sintéticos. Ante ello, se mencionan modelos específicos, como Med-PaLM y BioGPT, destinados a mejorar la toma de decisiones clínicas y la investigación biomédica. Finalmente, se constata que la GAI se utiliza para optimizar medicamentos, realizar simulaciones médicas, desarrollar chatbots para atención al paciente e identificar biomarcadores (Sai et al., 2024).

3.2 Los sesgos y sus implicancias en los resultados derivados de la aplicación de IA sobre datos de salud

En relación con los sesgos de la inteligencia artificial en el ámbito de la salud, esta revisión se enfoca en destacar las aportaciones de los autores al respecto, evidenciando los tipos de sesgos que identifican y las definiciones o interpretaciones que ofrecen sobre estos.

En principio, los autores refieren que, dentro de los problemas más relevantes que enfrentan las aplicaciones de IA en salud, están los del sesgo, pues los modelos de IA pueden verse significativamente afectados por datos incompletos o sesgados (Burlina et al., 2021; García-Gómez et al., 2024; Mascarenhas et al., 2023, 2024; Paik et al., 2023; Reina Reina et al., 2022; Rueda et al., 2024; Sai et al., 2024) o producir sesgos con sus resultados.

En general, los autores han definido ampliamente el sesgo como una disparidad en el tratamiento que se otorga en la atención médica debido al uso de IA. En ese sentido, mencionan que las disparidades pueden introducirse durante el entrenamiento de los modelos y su diseño.

Por otra parte, Paik et al. (2023) mencionan que la forma en la que se ha desarrollado la atención médica comúnmente a través del tiempo está intrínsecamente relacionada con los sesgos y, a menudo, se manifiesta sin que se perciba. De esta forma, aunque los profesionales tratan de evitar sesgos explícitos, los implícitos y sistémicos están presentes en las instituciones, permeando el desarrollo de la IA en ese ámbito. Aún más, Wilson et al. (2023) mencionan que la IA frecuentemente está influenciada por privilegios políticos, raciales y coloniales. Por consiguiente, estas influencias pueden crear sesgos en los sistemas de codificación de los algoritmos, lo que hace que se reflejen y se perpetúen las estructuras dominantes de nuestra sociedad. Dichas desigualdades sistemáticas que están arraigadas en las estructuras sociales, políticas y económicas (sesgo estructural) pueden manifestarse en la forma en que se diseñan y aplican los algoritmos de IA, produciendo decisiones clínicas que perduran o amplifican las disparidades existentes en la atención sanitaria (Aranovich y Matulionyte, 2023; Paik et al., 2023; Rueda et al., 2024; Wilson et al., 2023).

Por su parte, Aranovich y Matulionyte (2023) destacan que el sesgo estructural, incluyendo la discriminación contra minorías, es un problema identificado en muchas aplicaciones de salud basadas en IA. Este sesgo es particularmente significativo, ya que puede resultar en diagnósticos imprecisos y en la insuficiencia de atención adecuada a grupos subrepresentados (Fayos de Arizón et al., 2023). Por ejemplo, un estudio sobre la precisión de los oxímetros de pulso

reveló que en individuos con piel más oscura se tiende a sobreestimar la saturación de oxígeno arterial, lo que genera discrepancias en las intervenciones de tratamiento. Asimismo, otro estudio sobre los resultados de la artroplastia total de cadera evidenció que, un año después del procedimiento, las mujeres tenían una mayor probabilidad que los hombres de reportar la necesidad de asistencia en las actividades diarias (Paik et al., 2023).

Respecto a la calidad de los datos, es relevante destacar que los sistemas de aprendizaje automático se configuran según los datos con los que son entrenados y validados, identificando patrones en grandes conjuntos de datos que generan los resultados deseados, por tanto, su efectividad depende de las características de los datos en los que se basan (Mascarenhas et al., 2023). Al respecto, se ha informado sobre el sesgo de subrepresentación, que se refiere a la situación en la que ciertos grupos desfavorecidos o minoritarios están insuficientemente representados en los conjuntos de datos utilizados para entrenar modelos de IA. En este ámbito, las decisiones clínicas sesgadas y la perpetuación de desigualdades en la atención médica se producen porque los modelos pueden no ser capaces de generalizar adecuadamente a estas poblaciones subrepresentadas. Justamente, el aumento significativo que han tenido los datos de salud no se distribuyen de manera equitativa, ya que la mayor parte de los datos digitales disponibles provienen de áreas con mayores recursos, dejando fuera a importantes segmentos de la población Paik et al. (2023). Entre sus consecuencias, se destaca que los algoritmos de IA pueden ofrecer resultados menos precisos y equitativos, lo que afecta la calidad de la atención que reciben estos grupos (García-Gómez et al., 2024; Mascarenhas et al., 2023, 2024; Paik et al., 2023; Rueda et al., 2024; Sai et al., 2024).

Igualmente, muchas aplicaciones de GAI recopilan datos de internet sin identificar fuentes de desinformación. Por ejemplo, los *chatbots* que utilizan modelos de lenguaje de gran escala parecen creativos, pero solo reflejan patrones lingüísticos y no comprenden el lenguaje como lo hacen los humanos, amplificando contenidos dañinos, sin la capacidad de evaluar su naturaleza antes de responder a una consulta, como se ha evidenciado en múltiples casos (Sai et al., 2024).

De igual forma, la literatura analizada describe la existencia del sesgo de espectro, que se refiere a la situación en la que un disposi-

tivo de diagnóstico se aplica a una población que no corresponde al grupo para el cual fue diseñado. Esto puede afectar la precisión de los resultados, ya que el modelo pierde su efectividad en contextos diferentes a los que fue entrenado. Este tipo de sesgo tiene consecuencias en la implementación de herramientas de IA en la atención médica, ya que puede llevar a decisiones clínicas incorrectas y a una atención desigual (García-Gómez et al., 2024; Mascarenhas et al., 2023, 2024; Sai et al., 2024). Por ejemplo, este tipo de sesgo se ha identificado como un posible obstáculo para las aplicaciones de IA en la endoscopia cápsula, así como en el ámbito de la medicina cardiovascular (Mascarenhas et al., 2023).

En este contexto, también se ha mencionado el fenómeno de sobreajuste de los datos, que se presenta cuando un modelo de IA está demasiado ajustado a los datos de entrenamiento, lo que significa que ha aprendido patrones específicos de esos datos en lugar de generalizar a partir de ellos. Como resultado, el modelo puede funcionar muy bien con los datos de entrenamiento, pero su rendimiento se deteriora significativamente cuando se aplica a nuevos conjuntos de datos o a situaciones no vistas. Esto limita la aplicabilidad del modelo en entornos del mundo real, donde los datos pueden variar (Mascarenhas et al., 2023, 2024).

Finalmente, los autores destacan el "problema de atribución" en modelos de inteligencia artificial generativa en salud, que se refiere a la dificultad de explicar las decisiones tomadas por estos sistemas (Sai et al., 2024). Dicha opacidad o falta de transparencia en los algoritmos de IA puede generar desconfianza entre clínicos y pacientes; el fenómeno se describe como una "caja negra", donde los procesos internos del algoritmo no son accesibles ni comprensibles para los usuarios (García-Gómez et al., 2024; Marques et al., 2024; Rueda et al., 2024).

3.3 Síntesis sobre la aplicación de la IA en salud y los sesgos asociados

El análisis que se presenta en este apartado muestra claramente que, aunque la inteligencia artificial (IA) ha traído innovaciones significativas en varios ámbitos de la salud, su uso está marcado por diferen-

tes tipos de sesgos que afectan sus resultados. Estos sesgos, que han sido identificados en diversos estudios, se manifiestan en múltiples etapas del diseño, entrenamiento y aplicación de los sistemas, generando implicaciones clínicas, éticas y sociales.

Al respecto, se puede identificar el sesgo estructural, que se refiere a las desigualdades históricas y sistémicas que están profundamente arraigadas en nuestras estructuras sociales, políticas y económicas. Este tipo de sesgo muestra cómo la IA puede perpetuar privilegios raciales, de clase o de género que ya existen, ya que los sistemas se desarrollan dentro de instituciones que están influenciadas por estas dinámicas (Aranovich y Matulionyte, 2023; Wilson et al., 2023).

Por otro lado, el sesgo de subrepresentación aparece cuando ciertos grupos poblacionales —como minorías étnicas, mujeres o personas en situaciones de pobreza— están poco representados en los conjuntos de datos que alimentan los modelos. Esto impide que la IA pueda generalizar sus resultados de manera adecuada, lo que puede llevar a diagnósticos erróneos o tratamientos inadecuados para estos grupos (Mascarenhas et al., 2023; Paik et al., 2023).

En otro sentido, un sesgo técnico importante a considerar es el sobreajuste del modelo a los datos de entrenamiento, pues el modelo aprendió en exceso de los patrones específicos de esos datos. Lo anterior limita su capacidad para adaptarse a contextos clínicos nuevos o diferentes.

Además, se ha identificado el sesgo de espectro, que ocurre cuando un modelo se aplica fuera del contexto poblacional para el que fue diseñado, lo que reduce su precisión diagnóstica o terapéutica pues tiene limitada capacidad de responder en contextos con alta diversidad (Mascarenhas et al., 2024).

En otro sentido, en el ámbito de la inteligencia artificial generativa, hay un riesgo adicional relacionado con la recopilación masiva y sin criterio de datos de internet, que puede incluir información errónea, sesgada o perjudicial, sin contar con mecanismos efectivos para filtrar o contextualizar esa información (Sai et al., 2024).

Por último, hay una consecuencia crítica que surge del uso de la IA en el ámbito de la salud, que es el problema de atribución, el cual también se conoce como el "efecto caja negra". Así pues, aunque no se considera un sesgo en sí mismo, el problema de la atribución

aumenta los riesgos que conlleva el uso descontrolado de la IA en la salud y subraya la necesidad de crear modelos que sean más explicativos, auditables y en los que se pueda confiar.

Consecuentemente, los sesgos mencionados generan preocupaciones éticas y legales en la literatura sobre la industria de la salud. Por un lado, se plantea la responsabilidad legal por los daños causados por decisiones sesgadas (Sai et al., 2024) y, por otro, se enfatiza la necesidad de un desarrollo ético de las IA en salud (Mascarenhas et al., 2023). Probablemente, esto traería consigo la exigencia de mayor transparencia sobre las fuentes de datos utilizadas en el diseño y desarrollo de estos sistemas, con las consiguientes obligaciones de protección y seguridad de los datos (Mascarenhas et al., 2023).

4. Discusión

4.1 Evidencia del potencial transformador de la IA en salud y el impacto negativo de los sesgos

Los métodos de la revisión sistemática se fundamentan en la metodología PRISMA, que garantiza una evaluación rigurosa y clara de la literatura disponible. Se llevó a cabo una búsqueda exhaustiva en bases de datos relevantes y se aplicaron criterios de inclusión y exclusión para seleccionar estudios sobre el impacto de la IA en la atención médica. Este enfoque permitió identificar patrones y tendencias en la investigación, así como áreas que necesitan más estudio. En general, la evidencia muestra que la integración de IA en la atención médica ha generado un cambio paradigmático en la forma en que se diagnostican y tratan diversas enfermedades o trastornos de salud. Muchas de ellas tienen un enorme potencial para mejorar diagnósticos, tratamientos médicos y reducir los costos de estos (Garcia-Saiso et al., 2024; Mascarenhas et al., 2023, 2024; Rueda et al., 2024; Wilson et al., 2023). Desde ese punto de vista, en muchos casos se trata de un soporte positivo a las atenciones médicas y se debe seguir avanzando en investigaciones que permitan incrementar los beneficios de la IA en salud.

Como contrapartida, los artículos revisados han mostrado escasa o nula evidencia especialmente en los países latinoamericanos, donde el desarrollo académico en el tema sigue siendo limitado, lo

que ha restringido los resultados obtenidos. La mayoría de los textos analizados proviene de contextos europeos donde se han involucrado algunos investigadores de España o Portugal. Este problema está vinculado a la pobreza de datos en salud, ya que la mayoría de los estudios provienen de naciones desarrolladas, como Estados Unidos y varias naciones europeas. Para Paik et al. (2023), esta falta de representación es preocupante, dado que incluso en países ricos se observan disparidades significativas en los resultados de salud, impulsadas por factores socioeconómicos y raciales. Además, esta pobreza de datos se relaciona con el acceso inequitativo a internet y a tecnologías esenciales, ya que la exclusión de quienes no tienen acceso digital agrava las disparidades en la generación de datos y en el acceso a atención médica (Paik et al., 2023).

Paralelamente, los autores comparten algunos de los problemas que se podrían enfrentar al utilizar estas herramientas en la práctica clínica. En particular, Rueda et al. (2024) resaltan que los sesgos a menudo se consideran de manera peyorativa debido a que "nos alejan de la verdad y nos desvían de la justicia", provocando consecuencias clínicas injustificadas para ciertos pacientes. Paik et al. (2023) y Wilson et al. (2023) agregan que la existencia de sesgos provenientes de la calidad de los datos que alimentan a la IA perpetúan las desigualdades existentes en la atención médica. Es decir, los algoritmos de IA, a pesar de ser considerados "imparciales", pueden amplificar sesgos preexistentes, lo que puede resultar en exclusión social, diagnósticos erróneos y tratamientos inadecuados. En relación con ello, Reina et al. (2022) y Fayos de Arizón et al. (2023) resaltan las consecuencias negativas de entrenar estos algoritmos con datos poco representativos. Incluso, Mascarenhas et al. (2023) refuerzan el impacto negativo en la precisión diagnóstica producida por los sesgos que surgen de datos incompletos o malinterpretados.

Sumado a eso, la capacidad de generalización de los modelos de inteligencia generativa puede verse comprometida si no se actualizan periódicamente con datos recientes o si no logran adaptarse a las nuevas tendencias en el ámbito médico. Dichos modelos presentan una alta sensibilidad frente a la variabilidad y se requiere equilibrar la precisión del modelo con el cumplimiento de normativas éticas y de propiedad intelectual (Sai et al., 2024).

4.2 Aproximaciones generales sobre el papel de la regulación en la corrección de sesgos algorítmicos en salud

Los resultados de la revisión revelaron varios tipos de sesgos que pueden surgir de la aplicación de la IA, aunque sus consecuencias jurídicas se describen de manera general. Con relación a ello, es fundamental que los investigadores y desarrolladores consideren no solo la existencia de los sesgos, sino también su impacto. En esa línea, los autores aproximan sus conclusiones a sugerir regulaciones que permitan reducir los efectos negativos de los sesgos. Así, para algunos se requieren tanto marcos regulatorios robustos como una adecuada gestión de sesgos en los datos (Mascarenhas et al., 2024). En esa línea, Paik et al. (2023) indican que las regulaciones rigurosas permiten mitigar los sesgos y asegurar una atención equitativa en salud. Dichos marcos regulatorios serían la solución más efectiva para asegurar la calidad, la seguridad y la transparencia en el uso de la IA en la atención médica, pues así abordarían problemas de confianza y responsabilidad (Aranovich y Matulionyte, 2023). Sin embargo, la literatura no profundiza en dichos marcos regulatorios, en qué consisten y cuáles serían sus bases fundamentales.

Considerando los aspectos anteriores, este apartado tiene como objeto identificar algunas aproximaciones jurídicas generales que permitirían iniciar una discusión en orden a posibles resoluciones a algunos problemas derivados del uso de IA en salud, particularmente de los sesgos que producen.

En primer lugar, los sesgos y la discriminación algorítmica en el ámbito de la medicina pueden causar daños a las personas, lo que genera el problema de la responsabilidad civil por los efectos adversos ocasionados. En otras palabras, surge la cuestión de cómo se puede atribuir la responsabilidad a diferentes actores, incluidos los médicos que utilizan las recomendaciones basadas en inteligencia artificial, las instituciones de salud que optan por implementarlas y los desarrolladores de dichos sistemas. Según Price II et al. (2024), la responsabilidad médica se fundamenta en el deber de cumplir con un estándar de atención, el cual puede verse alterado por la incorporación de sistemas de IA. En cuanto a la responsabilidad institucional, los autores argumentan que podría establecerse mediante la teoría de la responsabilidad vicaria, lo que implicaría que un hospital sea

responsable por las decisiones de sus empleados que utilizan IA. Sin embargo, también observan que aún persisten ambigüedades respecto a la responsabilidad directa de las instituciones en este contexto, que puede depender de la supervisión que ejerzan y de la selección de las tecnologías implementadas. Por último, dado que el *software* a menudo se entiende como una herramienta auxiliar en la toma de decisiones clínicas, existe una dificultad para atribuir la responsabilidad a los desarrolladores de las tecnologías.

Al respecto, se analiza la legislación en el contexto estadounidense, donde se ha examinado cómo las leyes actuales no son suficientes para abordar la responsabilidad por discriminación algorítmica en el ámbito de la salud. Aunque se mencionan varias leyes federales que prohíben la discriminación en la atención médica, se argumenta que estas no se aplican adecuadamente a los desarrolladores de tecnología de salud y que muchas no cubren la "discriminación no intencional", que es común en el uso de algoritmos (Roberts y Salib, 2024). Por otra parte, se explica que la responsabilidad legal por discriminación en el ámbito de la atención médica se ve limitada por las protecciones actuales, que abarcan principalmente a proveedores de atención médica individuales e institucionales, pero no necesariamente a las entidades que desarrollan IA en salud (Roberts y Salib, 2024). En otro entorno normativo, podrían tener aplicación en la atribución de responsabilidad, dos directivas propuestas por la Comisión Europea en 2022: una sobre la responsabilidad por productos defectuosos y otra sobre la responsabilidad relacionada con IA. Estas directivas tienen como objetivo abordar la inadecuación de las reglas de responsabilidad actuales, especialmente aquellas basadas en la culpa, para gestionar las reclamaciones por daños causados por productos y servicios habilitados por IA (Price II et al., 2024).

En segundo lugar, dado que para abordar el problema del sesgo generado por el uso de métodos de IA en el ámbito de la salud, surge la pregunta de cómo puede el derecho promover la diversificación de los datos utilizados para entrenar estos modelos. En ese sentido, Roberts y Salib (2024) han formulado algunas propuestas de política que incentivan a las empresas e investigadores a recoger datos representativos, enfatizando la necesidad de recopilar datos que reflejen adecuadamente diversas poblaciones, especialmente aquellas

históricamente marginadas. Esto podría incluir la exigencia de que, tras la introducción de un producto al mercado, las empresas tengan un tiempo limitado para reunir datos que aseguren que el modelo funciona de manera efectiva para diversas poblaciones, o que se den incentivos a las empresas para diversificar los datos y penalizaciones para aquellos que no cumplan dentro de plazo. Asimismo, sugieren políticas que promueven el desarrollo de algoritmos correctivos que puedan ajustar las salidas de otros algoritmos, ayudando a corregir sesgos en las decisiones de salud, entre otras propuestas. En el contexto europeo, se ha mencionado que, pese a los avances en su regulación (como la creación de la Ley de IA, que es la legislación más significativa del mundo sobre IA), aún enfrentan desafíos significativos en cuanto a la diversidad, no discriminación y equidad en el uso de inteligencia artificial (IA) en salud. Dicho marco regulatorio, aunque exige que los datos sean representativos para prevenir sesgos y proteger a poblaciones subrepresentadas, no aborda esta necesidad de manera efectiva. Lo anterior promueve la existencia de un enfoque sistemático limitado y sin requisitos claros sobre la representación de datos, lo que reduce la efectividad de las disposiciones (Minssen et al., 2024).

En el ámbito latinoamericano, al discutir en general la discriminación algorítmica, se ha señalado que las respuestas regulatorias a menudo imitan enfoques hegemónicos y desestiman problemáticas locales. Por tanto, para abordar la discriminación algorítmica, es crucial integrar la digitalización con un activismo que promueva la diversidad en el desarrollo de algoritmos. En relación con ello, se requiere una regulación contextualizada de la IA que considere las particularidades culturales y sociales de la región (Smart, 2024).

Finalmente, dado que los problemas derivados de los sesgos y decisiones discriminatorias erosionan la confianza de los pacientes en el uso de la IA, se plantea la interrogante sobre qué medidas pueden fortalecer esa confianza, pues ello es esencial para proteger los derechos de los pacientes y asegurar que sus intereses sean considerados. En este contexto, es fundamental incorporar derechos humanos que contrarresten la discriminación y los sesgos en los algoritmos. Para algunos, es necesario avanzar hacia una nueva generación de derechos humanos que responda a los desafíos y a la necesidad de garantizar la dignidad, la equidad y la justicia (Sánchez Hernández,

2024). Pues, aunque la Constitución sigue vigente, su origen en una era predigital plantea interrogantes sobre su capacidad para abordar adecuadamente los retos de la discriminación algorítmica (Sánchez Hernández, 2024). Desde otra óptica, Celeste (2022) ha destacado que, desde la llegada de las tecnologías digitales, el número de violaciones a estos derechos ha aumentado, debido a los riesgos asociados con su uso extendido en diversos ámbitos. Como resultado, el equilibrio buscado por las normas constitucionales se ha visto alterado por la revolución digital, y el ecosistema constitucional debe reaccionar ante estos desafíos (Celeste, 2022). Por ejemplo, Solaiman (2025) postula que la actual ley de IA de la Unión Europea (AI Act) no aborda adecuadamente los aspectos éticos y las relaciones entre los diferentes actores involucrados, como pacientes, proveedores de salud, reguladores y desarrolladores de tecnología. Por tal motivo propone una *Bill of Rights* para la IA en el ámbito de la salud, que buscaría clarificar los derechos de los pacientes, como el derecho a la información transparente, a la protección de datos, a la explicabilidad de las decisiones de IA y a la rendición de cuentas en caso de daños, pues al enmarcar estas protecciones como derechos, se facilitaría la educación de los pacientes sobre lo que pueden esperar al interactuar con tecnologías de IA en salud.

Al respecto, la Organización Mundial de la Salud (2021) ha señalado que la comunidad internacional reconoce el potencial transformador de la IA y los grandes datos en el ámbito de la salud, particularmente en la mejora del acceso y la personalización de la atención. No obstante, se enfatiza la necesidad de un diseño e implementación responsables para mitigar los riesgos asociados a la deshumanización, la erosión de la autonomía individual y la vulneración de la privacidad del paciente. Se están adaptando y reevaluando los marcos legales existentes en materia de derechos humanos, bioética y privacidad para abordar específicamente los desafíos planteados por la IA, incluyendo la consideración del artículo 8 de la Convención Europea de Derechos Humanos y la Convención de Oviedo sobre los Derechos Humanos y la Biomedicina. A pesar de la existencia de estos marcos, se reconoce la necesidad de una mayor claridad en la aplicación de los estándares de derechos humanos a la IA para abordar la compleja interacción entre esta y los derechos fundamentales.

4.3 Consideraciones éticas aplicadas al uso de inteligencia artificial en la atención médica y sus sesgos

Los textos destacan ampliamente la necesidad de involucrar consideraciones éticas, pues se subraya la necesidad de utilizar conjuntos de datos representativos y de alta calidad para evitar sesgos que afecten la precisión diagnóstica y la equidad en la atención médica (Mascarenhas et al., 2024). En esa línea, Sai et al. (2024) han relevado que la construcción de marcos éticos aseguraría, entre otras cosas, la representación diversa de los datos, un punto también enfatizado por Garcia-Saiso et al. (2024). Igualmente, se ha dicho que se debe integrar la bioética en la implementación de tecnologías como la IA, aplicaciones móviles y telemedicina, con principios claros para abordar estos problemas y garantizar la justicia y la transparencia (Panadés Zafra et al., 2024). En este sentido, surge la pregunta: ¿qué principios deberían incorporarse para enfrentar estas dificultades? La respuesta no es del todo clara, sin embargo, autores como Nyrup y Cibralic (2024) destacan varios principios éticos relevantes en el contexto de la IA en la salud, sobre todo para hacer frente a las problemáticas que son objeto de este estudio.

En primer lugar, la equidad y la inclusión, ya que se busca que las aplicaciones de IA beneficien a todos, especialmente a grupos vulnerables. Esto implica considerar las desigualdades existentes y garantizar que las voces de las comunidades afectadas sean escuchadas. Otro principio clave es la transparencia y la rendición de cuentas, que exige que los sistemas de IA sean comprensibles y que sus decisiones sean explicables, además de que les permita a los profesionales de la salud y a los pacientes entender cómo se toman las decisiones. Por último, el texto aboga por un enfoque pragmático que fomente la participación democrática en la toma de decisiones, permitiendo la mejora moral continua y la adaptación de las tecnologías a las realidades locales. Estos principios son esenciales para abordar los desafíos éticos que plantea la implementación de la IA en el ámbito de la salud. Por otro lado, García-Gómez et al. (2024) han propuesto el concepto de un "pasaporte de IA" como una medida para mitigar los riesgos asociados con la IA en la atención médica. Este pasaporte se concibe como un documento vivo que detalla el propósito de cada sistema de IA, declaraciones éticas, información

sobre su entrenamiento, evaluación y posibles sesgos. Su objetivo es mejorar la transparencia y la seguridad en las aplicaciones médicas de la IA, sirviendo como un registro integral para pacientes, clínicos y otros interesados.

En este punto, es fundamental seguir realizando investigaciones que profundicen en la integración de la bioética en la implementación de tecnologías de salud, como la inteligencia artificial, para asegurar equidad y transparencia. Las diversas corrientes bioéticas ofrecen una rica variedad de perspectivas que deben explorarse para abordar los sesgos y desigualdades existentes. Así, se podrán establecer principios sólidos que guíen la práctica médica y tecnológica hacia un futuro más justo y accesible.

4.4 Desafíos de la propiedad intelectual, investigación y transferencia tecnológica frente a la transparencia algorítmica en salud

Durante esta investigación se han relevado diversos tipos de sesgos derivados del uso de la IA Sin embargo, también se ha abordado la opacidad de ciertos sistemas complejos en el contexto de la IA. Dicho efecto provoca que los usuarios, e incluso los desarrolladores del sistema, no entiendan completamente cómo se procesan los algoritmos, pues utilizan técnicas muy complejas que no son fácilmente interpretables (Aranovich y Matulionyte, 2023). Este fenómeno, conocido en la IA como "caja negra", plantea serios desafíos éticos y prácticos en el ámbito médico, lo que afecta la confianza en estos sistemas (Marques et al., 2024). Todo esto conlleva no solo riesgos exclusivamente técnicos, sino también aquellos que se relacionan estrechamente con los modelos de protección de la propiedad intelectual, las políticas públicas de ciencia y tecnología y los mecanismos de transferencia de tecnología.

En primer lugar, lo antedicho ha generado un debate respecto de aumentar la transparencia algorítmica con el objetivo de que sus procesos sean explicados de manera comprensible para las personas. En el ámbito de la salud, los médicos deberían ser capaces de comunicarles a los pacientes cómo se utiliza la IA en su atención, lo que podría tener implicaciones en la claridad del consentimiento infor-

mado otorgado por ellos (Solaiman y Cohen, 2024). Por ejemplo, el Reglamento Europeo de Protección de Datos no solo les otorga a las personas el derecho a conocer y cuestionar la información almacenada, sino que también establece el derecho a recibir una "explicación" sobre las decisiones tomadas mediante la creación de perfiles (Urueña, 2019). Igualmente, la AI Act europea establece requisitos específicos para la transparencia, especialmente en relación con los sistemas de IA de alto riesgo, que son aquellos que pueden tener un impacto significativo en la salud, la seguridad y los derechos fundamentales (Minssen et al., 2024). Sin embargo, estas regulaciones aún tienen limitaciones prácticas y se sugiere que futuras investigaciones y desarrollos se centren en mejorar la transparencia algorítmica (Minssen et al., 2024).

En segundo lugar, el sistema de propiedad intelectual, especialmente en lo que respecta a las patentes y la protección de secretos comerciales, presenta desafíos importantes en el ámbito de la inteligencia artificial aplicada a la salud. En especial, debido a que la posibilidad de proteger códigos y metodologías bajo estas figuras legales puede restringir el acceso público a información crucial, lo que complica la auditoría independiente y la validación de los resultados de los algoritmos. Este problema se vuelve aún más crítico en áreas sensibles como la salud pública, donde la transparencia y el acceso al conocimiento deberían ser prioritarios sobre los intereses comerciales. En este contexto, Cruz Rivero (2024) destaca la necesidad de que las instituciones desarrollen marcos de propiedad intelectual que logren un equilibrio adecuado entre la protección de derechos y la rendición de cuentas. Tschider y Ho (2024) añaden a esta discusión que, aunque la legislación de patentes exige una divulgación adecuada, muchas invenciones en IA no cumplen con este requisito en la práctica, lo que genera una falta de claridad sobre cómo funcionan realmente estos sistemas. Además, aunque la protección de secretos comerciales puede ser beneficiosa para las empresas, también puede dificultar la supervisión pública, poniendo en riesgo el interés general. Por esta razón, estos autores subrayan la necesidad urgente de encontrar un equilibrio entre la protección de la propiedad intelectual y la transparencia en el uso de la IA en el sector de la salud.

En tercer lugar, los sistemas de incentivos financieros y fiscales para

promover la investigación e innovación tecnológica —como las incubadoras de empresas, créditos preferenciales o subsidios— deben ir acompañados de salvaguardas éticas y regulatorias. Al respecto, si estos mecanismos de fomento no consideran criterios de inclusión y diversidad en el desarrollo de tecnologías de IA, se corre el riesgo de financiar soluciones que reproducen desigualdades estructurales. Así, la promoción de la investigación debe incluir no solo beneficios fiscales, sino también obligaciones vinculadas a la equidad, la transparencia de los modelos y la validación en contextos locales (Sai et al., 2024).

Por otro lado, la transferencia de tecnología en este ámbito a menudo se lleva a cabo sin un marco regulatorio sólido. La desregulación progresiva y la falta de supervisión sobre los contratos tecnológicos fomentan prácticas restrictivas que perjudican la libre competencia y el acceso equitativo a las innovaciones. Las cláusulas contractuales que limitan el uso o la interoperabilidad de herramientas basadas en IA, o que excluyen a ciertos actores del ecosistema de salud, pueden crear situaciones de dependencia tecnológica y agravar brechas ya existentes. Así, la carencia de una política clara sobre la transferencia tecnológica y la falta de control sobre cláusulas abusivas permiten que la innovación se enfoque más en la rentabilidad que en el bienestar común (Martínez Montenegro, 2022).

En suma, los sesgos en torno a la inteligencia artificial aplicada a la salud no son solo un problema de diseño técnico, sino que también son el resultado de estructuras institucionales, jurídicas y económicas que necesitan ser revisadas. Fomentar la innovación tecnológica con un enfoque inclusivo, proteger la propiedad intelectual sin obstaculizar la transparencia y regular adecuadamente la transferencia de tecnología son pasos fundamentales para que la IA realmente contribuya a sistemas de salud más justos, eficaces y humanos.

5. Conclusiones

Las tecnologías de la IA se presentan como potencial transformador en el sector salud, con la capacidad de mejorar diagnósticos, optimizar tratamientos y facilitar una atención médica más eficiente. Sin embargo, también plantean desafíos significativos, como la perpetuación de sesgos que afectan desproporcionadamente a grupos vul-

nerables, incluidos aquellos de minorías raciales y socioeconómicas. Estos sesgos tienen su origen principalmente en conjuntos de datos no representativos, lo que impacta negativamente en la precisión diagnóstica y exacerba las desigualdades existentes en la atención médica. Entre los tipos de sesgos identificados se destacan la subrepresentación de ciertos grupos poblacionales y el sesgo de espectro, ambos con implicaciones graves para la toma de decisiones clínicas.

La investigación subraya la necesidad de seguir analizando y desarrollando marcos regulatorios robustos que aborden los riesgos asociados con el uso de la IA en salud. Esto incluye establecer requisitos claros sobre la representación diversa de los datos y fomentar su diversificación para mitigar sesgos algorítmicos. Además, se destaca la importancia de avanzar en estudios relacionados con los aspectos éticos y jurídicos de estas tecnologías, especialmente en Latinoamérica, donde se observa un desarrollo limitado en comparación con países europeos pues dicha disparidad genera una brecha significativa en el conocimiento aplicable a contextos locales. En este caso, se reconoce la importancia de desarrollar regulaciones contextualizadas que integren las particularidades sociales, culturales y económicas propias de la región, de modo que puedan responder eficazmente a las problemáticas locales.

En concordancia con ello, la discusión también ofrece distintas propuestas que servirán de base para futuras investigaciones en el ámbito jurídico. Respecto a ello, y dado que aún persiste una considerable ambigüedad jurídica sobre la atribución de responsabilidad por la ocurrencia de daños derivados del uso de estas tecnologías, se proponen investigaciones que aborden el debate sobre si esta recae en los médicos que implementan las recomendaciones algorítmicas, en las instituciones de salud que adoptan estas herramientas o en los desarrolladores de los sistemas de IA.

Por otro lado, se proponen estudios sobre políticas públicas innovadoras que podrían incentivar la diversificación y representatividad de los datos utilizados para entrenar los modelos de IA, incluyendo mecanismos de sanciones y recompensas para fomentar la inclusión de poblaciones tradicionalmente marginadas, con el objetivo de reducir sesgos y mejorar la justicia en los resultados clínicos. Este problema está vinculado a la pobreza de datos en salud, lo cual es

preocupante dado que se ha mencionado que incluso en países ricos se observan disparidades significativas, impulsadas por factores socioeconómicos y raciales.

En consonancia con lo anterior, es preciso avanzar en la comprensión, análisis y mitigación del fenómeno conocido como la "caja negra". Lo anterior, porque la dificultad de entender y explicar cómo ciertos sistemas de IA procesan datos y generan decisiones produce desconfianza entre profesionales médicos, pacientes y reguladores. En esa línea, a pesar de que la legislación europea —como la AI Act— establece exigencias claras en materia de transparencia, aún falta evidencia que aborde las limitaciones prácticas que enfrentan las normas en sentido de lograr una explicabilidad completa y comprensible en contextos clínicos complejos.

En esa línea, futuras investigaciones podrían referirse a la tensión inherente entre la necesidad de transparencia y la protección de secretos comerciales, pues la confidencialidad empresarial restringe el acceso público a los detalles del funcionamiento interno de los algoritmos, lo que es necesario ponderar jurídicamente.

Paralelamente, dada la presencia de discriminaciones, exclusiones de acceso y falta de confianza de algunos pacientes en los sistemas de IA, es necesario seguir debatiendo sobre la necesidad de una nueva generación de derechos humanos digitales que respondan de manera adecuada a los desafíos planteados por el contexto tecnológico actual. Más aún, se propone atender la idea de adaptar marcos legales internacionales para que incorporen explícitamente las particularidades y riesgos asociados al uso de la IA.

En términos éticos, es fundamental garantizar la transparencia y rendición de cuentas en el uso de sistemas de IA en salud. Se propone el concepto innovador del "pasaporte de IA", un registro integral que permitiría documentar información sobre estos sistemas, mejorando su transparencia y seguridad. Asimismo, se enfatiza la necesidad de evolucionar el derecho constitucional hacia un "constitucionalismo digital" que adapte los valores fundamentales a las realidades tecnológicas emergentes. Este marco busca promover principios esenciales como la no discriminación, la privacidad y la transparencia para proteger derechos fundamentales en la era digital.

Finalmente, se concluye que, a medida que la IA continúa in-

tegrándose en el sector salud, es crucial abordar los desafíos éticos, legales y sociales que surgen. Esto incluye garantizar un consentimiento informado adecuado, equilibrar los derechos de pacientes con los intereses comerciales de desarrolladores tecnológicos y promover nuevos derechos humanos para contrarrestar la discriminación algorítmica. La implementación efectiva de estas medidas será esencial para asegurar una atención médica equitativa y justa en un contexto cada vez más digitalizado.

Bibliografía

Aranovich, T. d. C. y Matulionyte, R. (2023). Ensuring AI explainability in healthcare: problems and possible policy solutions. *Information & Communications Technology Law*, *32*(2), 259-275.

Azuaje Pirela, M. (2023). Propiedad intelectual como herramienta para promover la transparencia y prevenir la discriminación algorítmica. *Revista Chilena de Derecho y Tecnología*, *12*, 1-34. https://doi.org/10.5354/0719-2584.2023.70131

Breceda Pérez, J. A. y Castillo Lara, C. (2023). Derecho y ciencia: Entre la dignidad humana y la inteligencia artificial. *Ius et Scientia*, *9*(2), 261-287. https://doi.org/10.12795/IESTSCIENTIA.2023.i02.12

Burlina, P., Joshi, N., Paul, W., Pacheco, K. D. y Bressler, N. M. (2021). Addressing Artificial Intelligence Bias in Retinal Diagnostics. *Translational Vision Science y Technology*, *10*(2), 13-13.

Byrne, M. D. (2021). Reducing Bias in Healthcare Artificial Intelligence. *Journal of PeriAnesthesia Nursing*, *36*(3), 313-316.

Celeste, E. (2022). *Digital Constitutionalism: The Role of Internet Bills of Rights*. Editorial Milton: Taylor & Francis Group.

Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., Sahai, S. y Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering, 7*(6), 719-742. https://doi.org/10.1038/s41551-023-01056-8

Cosgrove, L., Karter, J. M., McGinley, M. y Morrill, Z. (2020). Digital Phenotyping and Digital Psychotropic Drugs: Mental Health Surveillance Tools That Threaten Human Rights. *Health and Human Rights*, *22*(2), 33-39.

Cruz Rivero, D. (2024). Disputabilidad y equidad en los mercados digitales: Una visión de Derecho europeo. *Revista Chilena de Derecho y Ciencia Política*, *15(1)*, 1-31. https://doi.org/10.7770/rchdcp-v15n1-art393

Davis, S. L. M. y Williams, C. (2020). Enter the cyborgs: Health and human rights in the digital age. *Health and Human Rights*, 22, 1-6.

Fayos de Arizón, L., Viera, E. R., Pilco, M., Perera, A., De Maeztu, G., Nicolau, A.,… Torra, R. (2023). Artificial intelligence: a new field of knowledge for nephrologists? *Clinical Kidney Journal, 16*(12), 2314-2326.

García Uribe, J. C. (2021). Propiedad intelectual, patentes y salud: Una mirada desde la bioética. *Revista Latinoamericana de Bioética*, *21*(2), 25-40. https://doi.org/10.18359/rlbi.5076

Garcia-Saiso, S., Marti, M., Pesce, K., Luciani, S., Mujica, O., Hennis, A. y D'Agostino, M. (2024). Artificial intelligence as a potential catalyst to a more equitable cancer care. *JMIR Cancer*, *10*, e57276. https://doi.org/10.2196/57276

García-Gómez, J. M., Blanes-Selva, V. y Doñate-Martínez, A. (2024). Proposing an AI Passport as a Mitigating Action of Risk Associated to Artificial Intelligence in Healthcare. *Stud Health Technol Inform*, *316*, 547-551. https://doi.org/10.3233/shti240472

Johnson, K. B., Wei, W. Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J. y Snowdon, J. L. (2021). Precision medicine, AI, and the future of personalized health care. *Clinical and Translational Science, 14*(1), 86-93. https://doi.org/10.1111/cts.12884

Price, W. N., Gerke, S. y Cohen, I. G. (2022). Liability for use of artificial intelligence in medicine. *Law & Economics Working Papers*, *241*. https://repository.law.umich.edu/law_econ_current/241

Lysaght, T., Lim, H. Y., Xafis, V. y Ngiam, K. Y. (2019). AI-Assisted Decision-making in Healthcare. *Asian Bioethics Review*, *11*(3), 299-314. https://doi.org/10.1007/s41649-019-00096-0

Marques, M., Almeida, A. y Pereira, H. (2024). The Medicine Revolution through Artificial Intelligence: Ethical Challenges of Machine Learning Algorithms in Decision-Making. *Cureus*, *16*(9), e69405. https://doi.org/10.7759/cureus.69405

Martínez Montenegro, I. (2022). *La importancia de resetear la cultura sociojurídica: Bases teóricas para la armonización e integración de lineamientos sociojurídicos. Novum Jus, 16*(3).

Martínez Montenegro, I. (2023). Sobre los métodos de la investigación jurídica. *Revista Chilena de Derecho y Ciencia Política, 14*(1), 1-4. https://doi.org/10.7770/rchdcp-v14n1-art312

Martínez Montenegro, I. (2024). *Propiedad intelectual y transferencia de tecnología en Chile*. Editorial Thomson Reuters.

Mascarenhas, M., Afonso, J., Ribeiro, T., Andrade, P., Cardoso, H. y Macedo, G. (2023). The Promise of Artificial Intelligence in Digestive Healthcare and the Bioethics Challenges It Presents. *Medicina (Kaunas)*, *59*(4). https://doi.org/10.3390/medicina59040790

Mascarenhas, M., Martins, M., Ribeiro, T., Afonso, J., Cardoso, P., Mendes, F.,… Macedo, G. (2024). Software as a Medical Device (SaMD) in Digestive Healthcare: Regulatory Challenges and Ethical Implications. *Diagnostics (Basel)*, *14*(18), 2100–2116. https://doi.org/10.3390/diagnostics14182100

Minssen, T., Solaiman, B., Wested, J., Köttering, L. y Malik, A. (2024). Governing AI in the European Union: Emerging infrastructures and regulatory ecosystems in health. En Solaiman, B. y Cohen, G. (Eds.), *Research handbook on health, AI and the law* (pp. 311-331). Edward Elgar Publishing. https://doi.org/10.4337/9781802205657.00027

Nyrup, R. y Cibralic, B. (2024). Idealism, realism, pragmatism: Three modes of theorising within secular AI ethics. En Solaiman, B. y Cohen, G. (Eds.), *Research handbook on health, AI and the law* (pp. 203-218). Edward Elgar Publishing.

Organización Mundial de la Salud. (2021a). *Estrategia mundial sobre salud digital 2020–2025 [Global strategy on digital health 2020–2025]*. https://www.who.int/publications/i/item/9789240020924

Organización Mundial de la Salud. (2021b). *Ética y gobernanza de la inteligencia artificial para la salud: Orientaciones de la OMS*. https://iris.who.int/handle/10665/341996

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P. y Moher, D. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *British Medical Journal*, (372), n160. https://doi.org/10.1136/bmj.n160

Paik, K. E., Hicklen, R., Kaggwa, F., Puyat, C. V., Nakayama, L. F., Ong, B. A., Walker, J. D., Shah, S. A., Laviada-Molina, H., Forde, K. A., Na, L., Bascunana, C., Onumah, C., Li, J. y Villanueva, C. (2023). Digital determinants of health: Health data poverty amplifies existing health disparities—A scoping review. *PLOS Digital Health*, *2*(10), e0000313. https://doi.org/10.1371/journal.pdig.0000313

Panadés Zafra, R., Amorós Parramon, N., Albiol-Perarnau, M. y Yuguero Torres, O. (2024). Análisis de retos y dilemas que deberá afrontar la bioética del siglo XXI, en la era de la salud digital. *Atención Primaria, 56*(7), 102901.

Price II, W. N., Gerke, S. y Cohen, I. G. (2024). Liability for use of artificial intelligence in medicine. En Solaiman, B. y Cohen, I. G. (Eds.), *Research handbook on health, AI and the law* (pp. 150-166). Edward Elgar Publishing. https://doi.org/10.4337/9781802205657.ch09

Ramón Fernández, F. (2021). *Inteligencia artificial en la relación médicopaciente: Algunas cuestiones y propuestas de mejora*. *Revista Chilena de Derecho y Tecnología, 10*(1), 329-351.

Reina Reina, A., Barrera, J. M., Valdivieso, B., Gas, M. E., Maté, A. y Trujillo, J. C. (2022). Machine learning model from a Spanish cohort for prediction of SARS-COV-2 mortality risk and critical patients. *Scientific Reports*, *12*(1), 5723. https://doi.org/10.1038/s41598-022-09613-y

Roberts, J. L. y Salib, P. (2024). Algorithmic discrimination and health equity. En Solaiman, B. y Cohen, I. G. (Eds.), *Research handbook on health, AI and the law* (pp. 93-110). Edward Elgar Publishing.

Rueda, J., Rodríguez, J. D., Jounou, I. P., Hortal-Carmona, J., Ausín, T. y Rodríguez-Arias, D. (2024). "Just" accuracy? Procedural fairness demands explainability in AI-based medical resource allocations. *AI & SOCIETY*, *39*(3), 1411-1422. https://doi.org/10.1007/s00146-022-01614-9

Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M. y Rodrigues, J. J. P. C. (2024). Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies, and Limitations. *IEEE Access*, *12*, 31078-31106.

Sekalala, S., Dagron, S., Forman, L. y Meier, B. M. (2020). Analyzing the Human Rights Impact of Increased Digital Public Health Surveillance during the COVID-19 Crisis. *Health and Human Rights*, *22*(2), 7-20.

Smart, S. (26 de agosto de 2024). *Algorithmic discrimination in Latin American welfare states*. *Harvard Kennedy School.*

Solaiman, B. (2025). The European Union's Artificial Intelligence Act and trust: Towards an AI Bill of Rights in healthcare? *Law, Innovation and Technology, 17*(1), 318-334. https://doi.org/10.1080/17579961.2025.2469986

Solaiman, B. y Cohen, I. G. (2024). A framework for health, AI and the law. En Solaiman, B. y Cohen, I. G. (Eds.), *Research handbook on health, AI and the law* (pp. 1-19). Edward Elgar Publishing.

Sun, N. N., Esom, K., Dhaliwal, M. y Amon, J. J. (2020). Human Rights and Digital Health Technologies. *Health and Human Rights*, *22*(2), 21-32.

Sánchez Hernández, J. (2024). La necesaria evolución de los derechos humanos: De su metamorfosis clásica hacia la presente encrucijada digital. *Revista de Derecho de la UNED*, *(33),* 617-637. https://doi.org/10.5944/rduned.33.2024.41940

Tschider, C. y Ho, C. M. (2024). Artificial Intelligence and Intellectual Property in Healthcare Technologies. En Solaiman, B. y Cohen, G. (Eds.), *Research handbook on health, AI and the law* (pp. 183-201). Edward Elgar Publishing.

Urueña, R. (2019). Autoridad algorítmica: ¿cómo empezar a pensar la protección de los derechos humanos en la era del "big data"? *Latin American Law Review*, *1*(2), 99-124.

Viniegra-Velázquez, L. (2024). El progreso en medicina y la inteligencia artificial. *Boletín Médico del Hospital Infantil de México*, *81*(3), 121-131. https://doi.org/10.24875/bmhim.24000007

Wilson, R. L., Higgins, O., Atem, J., Donaldson, A. E., Gildberg, F. A., Hooper, M. y Welsh, B. (2023). Artificial intelligence: An eye cast towards the mental health nursing horizon. *International Journal of Mental Health Nursing, 32*(3), 938-944. https://doi.org/10.1111/inm.13121

* * * *

Roles de autoría y conflicto de intereses

Los autores manifiestan haber cumplido los siguientes roles de autoría: **Bárcenas, J.:** concepción de la idea, metodología, conceptualización, visualización, escritura, revisión y edición; **Fuertes, C.:** escritura, revisión y edición; **Martínez Montenegro, I.:** conceptualización, escritura, revisión y edición.

Los autores manifiestan no poseer conflicto de interés alguno.

DOI: https://doi.org/10.26422/RIPI.2025.2200.bar