

# **Transferencia de sesgos y sesgo de automatización en radiología con inteligencia artificial**

Johan Flórez Vargas

Arturo Andrés Largo Navarro

Lizeth Carolina Ospina Gómez

Andrea Marcela Rodas Gallego

Asesor

Alberto Guzmán Avilés

Universidad Nacional Abierta y a Distancia – UNAD

Escuela de Ciencias de la Salud – ECISA

Tecnología en Radiología e Imágenes Diagnóstica

2026

## Resumen

El presente trabajo analiza el impacto de la inteligencia artificial (IA) en el campo de la radiología, centrándose en dos problemáticas relevantes: la transferencia de sesgos y el sesgo de automatización. La incorporación de sistemas basados en inteligencia artificial ha permitido mejorar la eficiencia en el análisis de imágenes médicas, optimizar la detección temprana de patologías y apoyar los procesos de toma de decisiones clínicas. Sin embargo, su implementación también plantea desafíos relacionados con la calidad de los datos utilizados para entrenar los algoritmos y con la confianza que los profesionales pueden depositar en las recomendaciones generadas por estos sistemas.

El objetivo de esta investigación es analizar cómo la transferencia de sesgos y el sesgo de automatización pueden influir en la práctica radiológica y afectar la precisión diagnóstica. Para ello, se desarrolló una investigación con enfoque cualitativo, de tipo descriptivo y documental, basada en la revisión de literatura científica relacionada con inteligencia artificial, radiología, sesgos algorítmicos y ética en salud. Los resultados evidencian que la transferencia de sesgos puede producirse cuando los algoritmos son entrenados con datos poco representativos o con errores, generando diferencias en el rendimiento diagnóstico entre diversos grupos poblacionales. Asimismo, el sesgo de automatización puede llevar a que los profesionales de la salud otorguen una confianza excesiva a las recomendaciones emitidas por los sistemas de inteligencia artificial, disminuyendo el análisis crítico y aumentando el riesgo de errores diagnósticos.

Se concluye que la inteligencia artificial debe ser utilizada como una herramienta de apoyo clínico que complemente el juicio profesional del radiólogo.

**Palabras Clave:** inteligencia artificial, radiología, transferencia de sesgos, sesgo de automatización, apoyo clínico, diagnóstico por imágenes.

## Abstract

This paper analyzes the impact of artificial intelligence (AI) in the field of radiology, focusing on two relevant issues: bias transfer and automation bias. The incorporation of AI-based systems has improved the efficiency of medical image analysis, optimized the early detection of pathologies, and supported clinical decision-making processes. However, their implementation also poses challenges related to the quality of the data used to train the algorithms and the trust that professionals can place in the recommendations generated by these systems.

The objective of this research is to analyze how bias transfer and automation bias can influence radiological practice and affect diagnostic accuracy. To this end, a qualitative, descriptive, and documentary research study was conducted, based on a review of scientific literature related to artificial intelligence, radiology, algorithmic bias, and ethics in healthcare. The results show that bias transfer can occur when algorithms are trained with unrepresentative data or with errors, generating differences in diagnostic performance among diverse population groups. Furthermore, automation bias can lead healthcare professionals to place excessive trust in the recommendations issued by artificial intelligence systems, diminishing critical analysis and increasing the risk of diagnostic errors.

It is concluded that artificial intelligence should be used as a clinical support tool that complements the radiologist's professional judgment.

**Keywords:** artificial intelligence, radiology, bias transfer, automation bias, clinical support, diagnostic imaging.

## Tabla de Contenido

Introducción .....	6
Planteamiento del Problema.....	8
Justificación.....	10
Objetivos .....	12
Objetivo General .....	12
Objetivos Específicos .....	12
Marco Teórico.....	13
Inteligencia Artificial en la Salud.....	13
Evolución de los Sistemas de Soporte a la Decisión Clínica (CDSS).....	15
Aprendizaje Automático (Machine Learning) .....	15
Deep Learning y Redes Neuronales Profundas.....	16
Implicaciones Éticas y Clínicas de la Inteligencia Artificial en Radiología.....	24
Evolución del Papel del Radiólogo en la Era de la Inteligencia Artificial.....	25
Comparación Internacional: Estados Unidos, Europa, Reino Unido, China y Colombia. ....	26
Estrategias de Mitigación de Sesgos y Riesgos Diagnósticos.....	27
Síntesis del Marco Teórico.....	28
Metodología .....	30
Criterios de Inclusión y Exclusión .....	31
<i>Criterios de Inclusión</i> .....	31
<i>Criterios de Exclusión</i> .....	31
Fases de la Investigación .....	32
<i>Fase 1. Búsqueda y Recopilación de la Información</i> .....	32

<i>Fase 2. Selección y Organización de la Información</i> .....	32
<i>Fase 3. Análisis e Interpretación de la Información</i> .....	32
<i>Fase 4. Elaboración del Documento Final</i> .....	33
Resultados.....	34
El Sesgo de Automatización: Cuando el Algoritmo Reemplaza al Juicio Clínico .....	34
La Transferencia de Sesgo: Cuando los Datos Reproducen Inequidades .....	35
Implicaciones éticas: Autonomía del Paciente y Justicia Algorítmica.....	37
Colaboración Interdisciplinaria: Un Requisito Ineludible.....	38
Desafíos Prácticos: Obstáculos Regulatorios, Económicos y de Carga Laboral .....	39
Evolución del Rol Profesional del Radiólogo .....	39
Marcos Regulatorios: Un Análisis Comparativo Internacional y el Contexto Latinoamericano.....	40
Implicaciones clínicas y estrategias de mitigación.....	43
Conclusiones .....	45
Referencias Bibliográficas.....	47

## Introducción

En las últimas décadas, el avance tecnológico ha transformado profundamente el ejercicio de la medicina, y la radiología no es la excepción. La inteligencia artificial (IA) se ha consolidado como una herramienta innovadora dentro del campo de la salud, gracias a su capacidad para procesar grandes volúmenes de información, identificar patrones complejos y optimizar el análisis de imágenes médicas, elementos fundamentales para el diagnóstico oportuno y preciso de diversas patologías. Como señala Menchú López (2025), estas tecnologías han permitido fortalecer la toma de decisiones clínicas, mejorar la eficiencia en los servicios de salud y reducir significativamente los tiempos de interpretación, posicionándose como un apoyo esencial para los profesionales.

En este mismo sentido, Chacón, S. M. (2025). destaca que, en el ámbito específico de la radiología, la inteligencia artificial funciona como un sistema complementario que facilita la detección temprana de enfermedades y refuerza la precisión diagnóstica; sin embargo, su valor real radica en su capacidad para asistir y no para sustituir el juicio clínico humano. La incorporación de estas herramientas ha abierto nuevas posibilidades, pero también ha planteado interrogantes y desafíos relacionados con la calidad de los datos, la equidad y la interacción entre el profesional y la tecnología.

Uno de los principales retos identificados en la literatura especializada es la presencia de sesgos dentro de los sistemas algorítmicos. Según Nair y colaboradores (2022), estos sesgos surgen cuando los conjuntos de datos utilizados para entrenar a los modelos contienen errores, limitaciones o no representan adecuadamente la diversidad de la población a la cual se aplicarán. Esta situación da origen a fenómenos críticos como la transferencia de sesgos, donde las

desigualdades o errores históricos presentes en los datos se replican y amplifican en los resultados diagnósticos, y el sesgo de automatización.

Estos fenómenos no solo afectan la validez de los resultados, sino que también tienen implicaciones directas sobre la seguridad del paciente, la equidad en la atención y la ética profesional. Por ello, resulta necesario analizar de manera crítica la implementación de la inteligencia artificial en la radiología, comprendiendo sus alcances, limitaciones y riesgos. El presente trabajo aborda en profundidad el estudio de la transferencia de sesgos y el sesgo de automatización, integrando elementos teóricos, metodológicos y normativos, con el propósito de promover un uso responsable, ético y seguro de estas tecnologías, garantizando que la innovación se traduzca siempre en una mejor calidad de atención para todos los pacientes.

## Planteamiento del Problema

La inteligencia artificial (IA) ha transformado significativamente el campo de la radiología al ofrecer herramientas capaces de analizar grandes volúmenes de imágenes médicas en tiempos reducidos y con altos niveles de precisión. Estas tecnologías han demostrado ser útiles para la detección temprana de enfermedades, la clasificación de hallazgos radiológicos y el apoyo a la toma de decisiones clínicas. Sin embargo, el creciente uso de sistemas basados en inteligencia artificial también ha generado preocupaciones relacionadas con la confiabilidad de los resultados y los riesgos asociados a la presencia de sesgos algorítmicos (Hosny et al., 2018).

Uno de los principales desafíos es la transferencia de sesgos, fenómeno que ocurre cuando los algoritmos aprenden patrones presentes en bases de datos que no representan adecuadamente la diversidad de la población. Como consecuencia, los sistemas pueden presentar un desempeño desigual entre distintos grupos demográficos, afectando la precisión diagnóstica y aumentando el riesgo de errores clínicos. Diversas investigaciones han evidenciado que algunos modelos de inteligencia artificial utilizados en imágenes médicas presentan diferencias en su rendimiento según variables como sexo, edad, origen étnico o contexto geográfico de los pacientes (Gichoya et al., 2023).

De igual manera, el sesgo de automatización representa una preocupación creciente en los entornos clínicos. Este fenómeno se presenta cuando los profesionales de la salud depositan una confianza excesiva en las recomendaciones generadas por sistemas automatizados, reduciendo la evaluación crítica de los resultados obtenidos. En radiología, esta situación puede llevar a aceptar interpretaciones incorrectas emitidas por la inteligencia artificial, incrementando la posibilidad de diagnósticos erróneos y afectando la seguridad del paciente (Dratsch et al., 2023).

La problemática adquiere una mayor relevancia en países de América Latina, donde gran parte de las herramientas de inteligencia artificial han sido desarrolladas y entrenadas con datos provenientes de otras regiones del mundo. Esto puede generar limitaciones en la aplicabilidad de los algoritmos a poblaciones con características epidemiológicas, sociales y demográficas diferentes. En Colombia, aunque se han impulsado iniciativas relacionadas con la transformación digital en salud, todavía existen desafíos asociados con la validación clínica de estas tecnologías, la regulación de su uso y la identificación de posibles sesgos en los sistemas implementados.

Ante este panorama, resulta necesario analizar los riesgos asociados al uso de la inteligencia artificial en radiología, especialmente aquellos relacionados con la transferencia de sesgos y el sesgo de automatización, debido a que pueden afectar la precisión diagnóstica, la equidad en la atención médica y la seguridad del paciente. Asimismo, es fundamental comprender el papel que desempeña el profesional de radiología en la supervisión y validación de los resultados generados por estas herramientas tecnológicas. En este contexto surge la siguiente pregunta de investigación: ¿De qué manera influyen la transferencia de sesgos y el sesgo de automatización en el uso de herramientas de inteligencia artificial para el apoyo en la toma de decisiones en radiología?

## **Justificación**

El estudio de la transferencia de sesgos y del sesgo de automatización en radiología resulta relevante debido al creciente uso de herramientas de inteligencia artificial en los procesos de diagnóstico médico. Estas tecnologías han permitido mejorar la eficiencia y precisión en el análisis de imágenes médicas, contribuyendo a la detección temprana de enfermedades, la optimización de los tiempos de respuesta y el fortalecimiento de los servicios de salud (Pesapane et al., 2018).

Sin embargo, el avance de la inteligencia artificial también ha generado nuevos desafíos relacionados con la calidad de los datos utilizados para entrenar los algoritmos y con la interacción de los profesionales de la salud frente a los sistemas automatizados. La presencia de sesgos en los modelos de inteligencia artificial puede generar diferencias en el rendimiento diagnóstico entre distintos grupos poblacionales, afectando la equidad en la atención médica y aumentando el riesgo de errores clínicos. Asimismo, el sesgo de automatización puede llevar a que los profesionales depositen una confianza excesiva en las recomendaciones emitidas por estas herramientas, disminuyendo el análisis crítico y el juicio clínico independiente (Geis et al., 2019; Cabitza et al., 2017).

Desde una perspectiva científica, esta investigación contribuye a la comprensión de los riesgos asociados a la implementación de sistemas de inteligencia artificial en radiología, permitiendo identificar factores que pueden afectar la precisión diagnóstica y la seguridad del paciente. Además, aporta elementos de análisis sobre la necesidad de desarrollar algoritmos más transparentes, representativos y confiables para su aplicación en contextos clínicos reales.

En el ámbito social y sanitario, el estudio adquiere relevancia debido a que las decisiones tomadas a partir de imágenes diagnósticas tienen un impacto directo en la salud y calidad de vida

de los pacientes. Un diagnóstico inexacto puede retrasar tratamientos, generar procedimientos innecesarios o afectar negativamente el pronóstico de una enfermedad. Por esta razón, resulta fundamental promover el uso ético y responsable de la inteligencia artificial, garantizando que estas herramientas contribuyan al bienestar de la población sin generar nuevas formas de desigualdad o discriminación (World Health Organization, 2021).

Desde el contexto latinoamericano y colombiano, la investigación cobra especial importancia debido a que gran parte de las herramientas de inteligencia artificial utilizadas en salud han sido desarrolladas con datos provenientes de otros países, lo que puede limitar su desempeño en poblaciones con características epidemiológicas, sociales y demográficas diferentes. Esto hace necesario fortalecer los procesos de validación clínica, regulación y supervisión de estas tecnologías antes de su implementación masiva en los servicios de salud.

Finalmente, la comprensión de estas tecnologías resulta fundamental para los profesionales de Radiología e Imágenes Diagnósticas, ya que fortalece el criterio clínico y promueve una práctica segura y responsable. De esta manera, la investigación contribuye tanto al desarrollo académico como al fortalecimiento de las competencias profesionales requeridas en los escenarios actuales de la atención en salud (European Society of Radiology, 2019).

## **Objetivos**

### **Objetivo General**

Desarrollar una revisión bibliográfica acerca del impacto de la transferencia de sesgos y sesgos y sesgos de automatización en la utilización de herramientas de inteligencia artificial aplicadas a la radiología describiendo sus implicaciones en la toma de decisiones clínicas y en la seguridad del paciente.

### **Objetivos Específicos**

Identificar los principales tipos de sesgos presentes en los sistemas de inteligencia artificial aplicados a la radiología y sus posibles efectos en la precisión diagnóstica.

Analizar la influencia del sesgo de automatización en el juicio clínico y en la toma de decisiones de los profesionales de la salud durante la interpretación de imágenes médicas.

Examinar el impacto de la transferencia de sesgos en el desempeño de las herramientas de inteligencia artificial utilizadas en radiología.

Comparar los avances, regulaciones y desafíos relacionados con la implementación de inteligencia artificial en radiología entre Colombia y otros contextos internacionales.

## Marco Teórico

### Inteligencia Artificial en la Salud

La inteligencia artificial (IA) se ha convertido en una de las innovaciones tecnológicas más importantes del siglo XXI debido a su capacidad para procesar grandes volúmenes de información, identificar patrones complejos y generar predicciones que apoyan la toma de decisiones. En el ámbito de la salud, estas tecnologías han adquirido una relevancia creciente porque permiten optimizar procesos clínicos, mejorar la precisión diagnóstica y contribuir a una atención más eficiente y personalizada. Según Davenport y Kalakota (2019), la IA tiene el potencial de transformar significativamente los sistemas sanitarios al facilitar el análisis de datos clínicos que, por su volumen y complejidad, resultan difíciles de procesar mediante métodos tradicionales.

El interés por incorporar inteligencia artificial en los servicios de salud surge como respuesta a múltiples desafíos contemporáneos, entre ellos el incremento de enfermedades crónicas, el envejecimiento poblacional y la creciente demanda de servicios médicos especializados. En este contexto, la IA se presenta como una herramienta capaz de complementar el trabajo de los profesionales de la salud mediante el procesamiento rápido y preciso de información clínica. Sin embargo, su implementación también plantea interrogantes relacionados con la seguridad, la ética y la calidad de las decisiones clínicas que se toman con apoyo de sistemas automatizados.

Actualmente, las aplicaciones de inteligencia artificial abarcan múltiples áreas de la medicina, incluyendo la predicción de enfermedades, el monitoreo remoto de pacientes, la gestión hospitalaria, la medicina personalizada y el análisis de imágenes diagnósticas. Estas herramientas no buscan reemplazar al profesional de la salud, sino fortalecer su capacidad de

análisis y facilitar procesos que permitan una atención más segura y efectiva (Jha & Topol, 2016).

La radiología ha sido una de las especialidades médicas que más rápidamente ha incorporado herramientas de inteligencia artificial debido a la naturaleza digital de las imágenes médicas. La disponibilidad de grandes bases de datos de radiografías, tomografías computarizadas, resonancias magnéticas y mamografías ha permitido desarrollar algoritmos capaces de identificar patrones complejos y detectar alteraciones anatómicas con altos niveles de precisión.

Según Hosny et al. (2018), la inteligencia artificial aplicada a radiología puede mejorar significativamente la eficiencia diagnóstica mediante la automatización de tareas repetitivas, la priorización de estudios urgentes y la detección temprana de patologías. Estas capacidades son especialmente importantes en contextos donde existe una alta demanda de estudios imagenológicos y una disponibilidad limitada de especialistas.

Entre las aplicaciones más destacadas se encuentran la detección de nódulos pulmonares en tomografía computarizada, la identificación de fracturas en radiografías convencionales y el análisis automatizado de lesiones mamarias. Estas herramientas permiten reducir los tiempos de interpretación y apoyar la toma de decisiones clínicas, favoreciendo diagnósticos más oportunos.

No obstante, diversos autores coinciden en que la inteligencia artificial debe considerarse un sistema de apoyo y no un sustituto del criterio profesional. La interpretación de una imagen médica requiere integrar información clínica, antecedentes del paciente y contexto epidemiológico, elementos que aún dependen en gran medida del juicio humano (McBee et al., 2018).

## **Evolución de los Sistemas de Soporte a la Decisión Clínica (CDSS)**

Los Sistemas de Soporte a la Decisión Clínica (Clinical Decision Support Systems, CDSS) constituyen uno de los antecedentes más importantes de la inteligencia artificial aplicada a la salud. Estos sistemas fueron diseñados para proporcionar recomendaciones basadas en evidencia científica con el objetivo de apoyar a los profesionales durante la toma de decisiones diagnósticas y terapéuticas.

Las primeras generaciones de CDSS funcionaban mediante reglas predefinidas desarrolladas por expertos clínicos. Aunque estos sistemas ofrecían apoyo útil, presentaban limitaciones importantes debido a su incapacidad para adaptarse a situaciones clínicas complejas o variables no contempladas previamente. Con el avance de la informática médica y el desarrollo de nuevas técnicas computacionales, los CDSS evolucionaron hacia modelos más dinámicos capaces de aprender a partir de los datos.

Actualmente, los sistemas modernos incorporan algoritmos de aprendizaje automático y análisis predictivo que permiten procesar grandes cantidades de información clínica en tiempo real. En radiología, estos sistemas pueden identificar hallazgos sospechosos, sugerir diagnósticos diferenciales y priorizar estudios según niveles de urgencia, mejorando así la eficiencia operativa de los servicios de imágenes diagnósticas (Kohli et al., 2017).

A pesar de sus beneficios, la creciente dependencia de estos sistemas también ha generado preocupación respecto a fenómenos como el sesgo de automatización, donde los profesionales pueden confiar excesivamente en las recomendaciones emitidas por los algoritmos.

## **Aprendizaje Automático (Machine Learning)**

El aprendizaje automático o Machine Learning constituye una de las principales ramas de la inteligencia artificial. Su funcionamiento se basa en la capacidad de los sistemas informáticos

para identificar patrones dentro de los datos y utilizar esa información para generar predicciones o clasificaciones sin necesidad de programación específica para cada situación.

Beam y Kohane (2018) señalan que el aprendizaje automático permite desarrollar modelos capaces de mejorar progresivamente su rendimiento a medida que reciben nuevos datos. Esta característica ha impulsado su aplicación en numerosos campos de la medicina, especialmente en aquellos que requieren el análisis de grandes volúmenes de información.

En radiología, los algoritmos de Machine Learning son utilizados para clasificar imágenes, identificar anomalías y apoyar la detección de enfermedades. Estos sistemas aprenden a partir de miles de estudios previamente etiquetados por especialistas y posteriormente aplican ese conocimiento para analizar nuevas imágenes. Sin embargo, el desempeño de estos algoritmos depende directamente de la calidad de los datos utilizados durante su entrenamiento. Cuando los conjuntos de datos presentan errores, sesgos o falta de representatividad poblacional, existe el riesgo de generar resultados inexactos que pueden afectar la calidad de la atención médica y la seguridad del paciente.

### **Deep Learning y Redes Neuronales Profundas**

El Deep Learning o aprendizaje profundo representa una evolución del aprendizaje automático basada en redes neuronales artificiales inspiradas en el funcionamiento del cerebro humano. Estas redes están compuestas por múltiples capas de procesamiento que permiten identificar características cada vez más complejas dentro de una imagen médica.

De acuerdo con Litjens et al. (2017), el Deep Learning ha revolucionado el análisis de imágenes diagnósticas debido a su capacidad para aprender directamente de los datos sin requerir una intervención humana constante en la identificación de características relevantes. Esta tecnología ha demostrado resultados prometedores en áreas como la detección de cáncer de

mama, enfermedades pulmonares y trastornos neurológicos. Uno de los ejemplos más conocidos es el desarrollo de sistemas capaces de detectar neumonía en radiografías de tórax con niveles de precisión comparables a los de especialistas entrenados (Rajpurkar et al., 2017). Estos avances han generado grandes expectativas respecto al potencial de la inteligencia artificial como herramienta de apoyo clínico.

No obstante, el uso de Deep Learning también plantea desafíos importantes relacionados con la transparencia de los algoritmos, la interpretabilidad de los resultados y la presencia de sesgos en los datos de entrenamiento. Por esta razón, la supervisión humana continúa siendo un componente indispensable para garantizar diagnósticos seguros, éticos y confiables.

### **Modelo de Aceptación Tecnológica (TAM)**

La incorporación de nuevas tecnologías en el ámbito sanitario no depende únicamente de su disponibilidad o de sus capacidades técnicas. También está relacionada con la percepción que los profesionales tienen sobre su utilidad y facilidad de uso. En este contexto, el Modelo de Aceptación Tecnológica (Technology Acceptance Model, TAM), propuesto por Davis (1989), constituye una de las teorías más utilizadas para explicar por qué las personas aceptan o rechazan una determinada tecnología.

Este modelo plantea que existen dos factores fundamentales que influyen en la adopción tecnológica: la utilidad percibida y la facilidad de uso percibida. La utilidad percibida se refiere al grado en que una persona considera que una herramienta mejorará su desempeño laboral, mientras que la facilidad de uso percibida corresponde a la creencia de que la tecnología puede utilizarse sin requerir un esfuerzo excesivo.

En radiología, la aceptación de herramientas basadas en inteligencia artificial depende en gran medida de la confianza que los profesionales depositan en sus resultados. Cuando los

radiólogos perciben que estas tecnologías contribuyen a mejorar la precisión diagnóstica, optimizar los tiempos de trabajo y reducir errores, existe una mayor probabilidad de adopción dentro de la práctica clínica (Pinto Dos Santos et al., 2019).

Sin embargo, una aceptación excesivamente positiva también puede generar riesgos. Cuando los profesionales desarrollan una confianza desmedida en los sistemas automatizados, pueden disminuir su capacidad crítica frente a las recomendaciones emitidas por los algoritmos. Por esta razón, diversos autores señalan que la integración de la inteligencia artificial debe realizarse de manera equilibrada, promoviendo una relación de colaboración entre la tecnología y el juicio clínico humano.

### **Efecto de Anclaje**

El efecto de anclaje es un sesgo cognitivo ampliamente estudiado en psicología que ocurre cuando una persona basa sus decisiones en la primera información que recibe, otorgándole una importancia desproporcionada frente a datos posteriores. Este fenómeno influye en numerosos procesos de toma de decisiones y también puede presentarse en el ámbito sanitario (Croskerry, 2003). En radiología, el efecto de anclaje puede manifestarse cuando un sistema de inteligencia artificial proporciona una sugerencia diagnóstica inicial y el profesional orienta su interpretación a partir de dicha recomendación. Como consecuencia, existe el riesgo de que hallazgos adicionales o información clínica relevante reciban menor atención, afectando la calidad del análisis diagnóstico (Cabitza, Rasoini, & Gensini, 2017)..La presencia de este sesgo resulta especialmente preocupante cuando las recomendaciones emitidas por los sistemas automatizados son incorrectas. En estos casos, el profesional puede mantener una interpretación influenciada por la primera impresión proporcionada por la herramienta tecnológica, aun cuando existan evidencias que sugieran una conclusión diferente (Goddard, Roudsari, & Wyatt, 2012).

Por esta razón, diversos investigadores destacan la importancia de promover procesos de verificación independiente y razonamiento crítico durante la interpretación de imágenes médicas. La inteligencia artificial debe considerarse un apoyo para la toma de decisiones y no una fuente absoluta de verdad diagnóstica (Topol, 2019).

### **Teoría de Detección de Señales (Signal Detection Theory - SDT)**

La Teoría de Detección de Señales (Signal Detection Theory, SDT) constituye un modelo ampliamente utilizado para comprender cómo los individuos distinguen entre señales reales y ruido dentro de un proceso de observación. Esta teoría tiene una aplicación directa en radiología

debido a que los profesionales deben identificar hallazgos patológicos dentro de imágenes que frecuentemente contienen elementos anatómicos complejos y posibles distractores (Green & Swets, 1966). La SDT permite analizar conceptos fundamentales como sensibilidad, especificidad, falsos positivos y falsos negativos. La sensibilidad corresponde a la capacidad de detectar correctamente una enfermedad cuando esta está presente, mientras que la especificidad se relaciona con la capacidad de descartar correctamente una enfermedad cuando en realidad no existe (Macmillan & Creelman, 2005).

La inteligencia artificial ha demostrado mejorar la sensibilidad diagnóstica en diferentes contextos clínicos; sin embargo, también puede incrementar la aparición de falsos positivos cuando los algoritmos identifican como patológicos hallazgos que en realidad son normales. Del mismo modo, un modelo insuficientemente entrenado puede generar falsos negativos que retrasen el diagnóstico de una enfermedad (Topol, 2019). Desde esta perspectiva, la SDT permite comprender que la precisión diagnóstica no depende únicamente de la capacidad tecnológica del

sistema, sino también de la interacción entre la herramienta, el profesional y el contexto clínico en el que se desarrolla la toma de decisiones (Green & Swets, 1966).

### **Sesgos en la Inteligencia Artificial**

A pesar de los beneficios que ofrece la inteligencia artificial en el ámbito sanitario, su implementación no está exenta de limitaciones. Una de las principales preocupaciones actuales corresponde a la presencia de sesgos algorítmicos, entendidos como errores sistemáticos que afectan la forma en que los sistemas procesan la información y generan resultados (Mehrabi et al., 2021). Los sesgos pueden originarse en diferentes etapas del desarrollo de un algoritmo. Frecuentemente aparecen durante la recopilación de datos, cuando las bases de entrenamiento no representan adecuadamente la diversidad de la población. También pueden generarse durante el etiquetado de la información o como consecuencia de decisiones adoptadas durante el diseño y validación de los modelos (Parikh, Teeple, & Navathe, 2019). En radiología, estos sesgos pueden tener consecuencias importantes sobre la calidad diagnóstica. Un sistema entrenado principalmente con imágenes de determinadas poblaciones puede presentar un rendimiento inferior cuando se utiliza en pacientes con características demográficas diferentes (Seyyed-Kalantari et al., 2021).

Esta situación puede afectar la precisión diagnóstica y contribuir a la aparición de desigualdades en el acceso a una atención médica segura y efectiva (Gichoya et al., 2023). La identificación y corrección de los sesgos constituye actualmente uno de los principales desafíos para el desarrollo responsable de herramientas de inteligencia artificial aplicadas a la salud. Por esta razón, organismos internacionales, investigadores y profesionales sanitarios coinciden en la necesidad de promover modelos transparentes, éticos y representativos de la diversidad poblacional.

## **Transferencia de Sesgos**

La transferencia de sesgos constituye uno de los principales desafíos asociados al uso de inteligencia artificial en el ámbito sanitario. Este fenómeno ocurre cuando los errores, limitaciones o desequilibrios presentes en los datos utilizados para entrenar un algoritmo son reproducidos posteriormente durante su aplicación clínica. En otras palabras, el sistema aprende patrones sesgados y los replica en nuevas situaciones, generando resultados que pueden afectar la precisión diagnóstica y la calidad de la atención médica. En radiología, la transferencia de sesgos suele estar relacionada con la falta de representatividad de los conjuntos de datos utilizados durante el entrenamiento de los modelos. Cuando las imágenes provienen principalmente de una población específica, el sistema puede desarrollar un mejor desempeño para determinados grupos de pacientes y una menor precisión para otros. Esta situación es especialmente preocupante en contextos multiculturales donde existen diferencias demográficas, genéticas y epidemiológicas importantes.

Diversas investigaciones han demostrado que algunos algoritmos presentan variaciones en su rendimiento según factores como la edad, el sexo, el origen étnico o las condiciones socioeconómicas de los pacientes. Banerjee et al. (2023) identificaron que ciertos modelos de inteligencia artificial utilizados en imágenes médicas tendían a presentar menores niveles de precisión en grupos históricamente subrepresentados dentro de los conjuntos de entrenamiento.

Para países latinoamericanos como Colombia, esta problemática adquiere una relevancia especial debido a que gran parte de las herramientas de inteligencia artificial disponibles han sido desarrolladas y entrenadas con datos provenientes de Estados Unidos, Europa o Asia.

Como consecuencia, existe el riesgo de que estos sistemas no reflejen adecuadamente las características clínicas y demográficas de la población local, generando posibles errores diagnósticos o disminuyendo la calidad de las recomendaciones emitidas por el algoritmo.

La transferencia de sesgos no debe entenderse únicamente como una limitación técnica. También representa un desafío ético y social, ya que puede contribuir a reproducir desigualdades existentes dentro de los sistemas de salud. Por esta razón, los expertos recomiendan desarrollar modelos más representativos, incorporar mecanismos de auditoría permanente y validar las herramientas en diferentes contextos poblacionales antes de su implementación clínica.

### **Sesgo de Automatización**

El sesgo de automatización hace referencia a la tendencia de las personas a confiar excesivamente en los resultados generados por sistemas automatizados, disminuyendo su capacidad de análisis crítico frente a la información proporcionada por la tecnología.

Este fenómeno ha sido ampliamente estudiado en sectores como la aviación, la ingeniería y, más recientemente, en el ámbito de la salud. En radiología, el sesgo de automatización puede aparecer cuando el profesional acepta las recomendaciones emitidas por un sistema de inteligencia artificial sin realizar una evaluación independiente de las imágenes diagnósticas. Aunque estas herramientas ofrecen importantes beneficios en términos de eficiencia y apoyo clínico, la confianza excesiva en sus resultados puede incrementar el riesgo de errores diagnósticos. Dratsch et al. (2023) demostraron que los radiólogos pueden verse influenciados significativamente por sugerencias incorrectas generadas por sistemas automatizados.

En algunos casos, incluso especialistas con amplia experiencia modificaron sus interpretaciones iniciales debido a la confianza depositada en las recomendaciones del algoritmo.

Estos hallazgos evidencian que el sesgo de automatización no afecta únicamente a profesionales en formación, sino también a expertos con años de experiencia clínica.

La aparición de este fenómeno suele estar relacionada con factores como la carga laboral, la presión asistencial, la complejidad de los estudios diagnósticos y la percepción de confiabilidad de la tecnología. Cuando los sistemas de inteligencia artificial son presentados como herramientas altamente precisas, los profesionales pueden desarrollar una falsa sensación de seguridad que reduce la verificación independiente de los resultados. Por esta razón, diversos organismos internacionales han insistido en la necesidad de mantener la supervisión humana como elemento fundamental dentro del proceso diagnóstico. La inteligencia artificial debe entenderse como un recurso de apoyo que complementa el trabajo del profesional, pero nunca como un sustituto de su criterio clínico.

### **Inteligencia Artificial como Apoyo Clínico**

La creciente incorporación de inteligencia artificial en los servicios de salud ha generado debates sobre el papel que estas tecnologías desempeñarán en el futuro de la práctica médica. Aunque algunos sectores han planteado la posibilidad de una automatización progresiva de determinadas tareas diagnósticas, la evidencia científica actual sugiere que la inteligencia artificial funciona de manera más efectiva cuando se integra como una herramienta de apoyo clínico. Desde esta perspectiva, la IA permite optimizar procesos, agilizar la interpretación de imágenes y facilitar la identificación de hallazgos relevantes. Sin embargo, las decisiones finales continúan dependiendo del conocimiento, la experiencia y el juicio profesional de los especialistas. La combinación entre inteligencia artificial y criterio humano ofrece mayores niveles de precisión diagnóstica que cualquiera de estos elementos por separado.

Además de mejorar la eficiencia, la inteligencia artificial puede contribuir a disminuir la carga laboral de los servicios de radiología, permitiendo que los especialistas dediquen más tiempo a casos complejos, actividades de investigación y comunicación interdisciplinaria. Este enfoque favorece una atención centrada en el paciente y fortalece la calidad de los procesos clínicos.

No obstante, para que estos beneficios se materialicen, es indispensable garantizar una implementación responsable de la tecnología, acompañada de procesos de formación continua, supervisión ética y evaluación permanente de los resultados obtenidos en la práctica clínica.

### **Implicaciones Éticas y Clínicas de la Inteligencia Artificial en Radiología**

La implementación de sistemas de inteligencia artificial en radiología no solo plantea desafíos tecnológicos, sino también importantes implicaciones éticas y clínicas. Aunque estas herramientas ofrecen ventajas significativas en términos de rapidez y precisión diagnóstica, su uso debe estar acompañado de mecanismos que garanticen la seguridad del paciente, la transparencia de los algoritmos y la responsabilidad profesional (Geis et al., 2019; European Society of Radiology, 2019). Uno de los principales debates éticos está relacionado con la autonomía del paciente. En muchos casos, los pacientes desconocen que parte de su proceso diagnóstico puede estar siendo influenciado por herramientas de inteligencia artificial. Esta situación genera interrogantes sobre la necesidad de informar adecuadamente a los usuarios acerca del papel que desempeña la tecnología dentro de su atención médica Gichoya et (2023). Asimismo, surge la preocupación sobre la responsabilidad profesional cuando ocurre un error diagnóstico asociado al uso de sistemas automatizados. Aunque la inteligencia artificial puede emitir recomendaciones, la responsabilidad final continúa recayendo sobre el profesional de la salud encargado de interpretar los hallazgos y tomar decisiones clínicas. Por esta razón, las

organizaciones científicas coinciden en que la supervisión humana debe mantenerse como un elemento indispensable dentro de cualquier proceso asistido por inteligencia artificial (Geis et al., 2019). Otro aspecto relevante corresponde a la equidad en el acceso a la atención médica. Si los algoritmos presentan diferencias de rendimiento entre grupos poblacionales, existe el riesgo de incrementar desigualdades ya existentes dentro de los sistemas sanitarios. Por ello, el desarrollo de herramientas de inteligencia artificial debe orientarse hacia modelos inclusivos, transparentes y representativos de la diversidad de pacientes que conforman la población (Parikh et al., 2019).

### **Evolución del Papel del Radiólogo en la Era de la Inteligencia Artificial**

La incorporación progresiva de inteligencia artificial ha generado inquietudes respecto al futuro de la profesión radiológica. Durante los últimos años han surgido múltiples debates sobre la posibilidad de que los algoritmos sustituyan parte de las funciones realizadas tradicionalmente por los especialistas en imágenes diagnósticas. Sin embargo, la evidencia científica actual indica que la inteligencia artificial no reemplaza al radiólogo, sino que transforma y complementa sus funciones (Topol, 2019; Jha & Topol, 2016).

En este nuevo escenario, el profesional deja de concentrarse exclusivamente en la identificación de hallazgos radiológicos para asumir un papel más amplio relacionado con la validación de resultados, la integración de información clínica y la toma de decisiones complejas.

Las herramientas automatizadas pueden detectar patrones y generar alertas, pero carecen de la capacidad necesaria para interpretar completamente el contexto clínico de cada paciente (European Society of Radiology, 2019). Además, la evolución tecnológica exige nuevas competencias profesionales. Los radiólogos del futuro deberán comprender los principios básicos del funcionamiento de los algoritmos, identificar posibles limitaciones de los sistemas y

participar activamente en la evaluación crítica de las herramientas utilizadas en su práctica clínica. Esta alfabetización tecnológica permitirá minimizar riesgos asociados al sesgo de automatización y fortalecer la seguridad diagnóstica. En este mismo sentido, Chacón, S. M. (2025) destaca que, en el ámbito específico de la radiología, la inteligencia artificial funciona como un sistema complementario que facilita la detección temprana de enfermedades y refuerza la precisión diagnóstica; sin embargo, su valor real radica en su capacidad para asistir y no para sustituir el juicio clínico humano.

Por otra parte, la automatización de tareas repetitivas puede liberar tiempo para que los especialistas participen de manera más activa en procesos interdisciplinarios, actividades académicas e interacciones directas con pacientes y otros profesionales de la salud. De esta manera, la inteligencia artificial puede contribuir al fortalecimiento del rol clínico y consultivo del radiólogo dentro de los equipos asistenciales (Obermeyer et al., 2019).

### **Comparación Internacional: Estados Unidos, Europa, Reino Unido, China y Colombia**

La implementación de inteligencia artificial en radiología ha seguido trayectorias diferentes según las características regulatorias, tecnológicas y económicas de cada país. Estados Unidos es considerado uno de los principales referentes mundiales debido a la participación activa de instituciones académicas, hospitales universitarios y organismos reguladores como la Food and Drug Administration (FDA). Este organismo ha desarrollado marcos específicos para evaluar software médico basado en inteligencia artificial, promoviendo procesos de validación y seguimiento continuo de estas tecnologías (Banerjee et al., 2023). En Europa, la regulación ha avanzado significativamente mediante la aprobación del Reglamento Europeo de Inteligencia Artificial (AI Act), que clasifica las aplicaciones médicas de IA como sistemas de alto riesgo. Esta normativa establece requisitos relacionados con transparencia, supervisión humana,

trazabilidad y protección de los derechos fundamentales de los pacientes. El objetivo principal consiste en garantizar que la innovación tecnológica se desarrolle bajo principios éticos y de seguridad (European Commission, 2024). El Reino Unido también ha impulsado estrategias orientadas a la evaluación de tecnologías sanitarias basadas en inteligencia artificial. A través del National Health Service (NHS) y del National Institute for Health and Care Excellence (NICE), se han promovido estándares para la validación clínica y la implementación segura de algoritmos diagnósticos dentro de los servicios de salud (NICE, 2022).

Por su parte, China se ha convertido en uno de los líderes mundiales en investigación y desarrollo de inteligencia artificial gracias a la disponibilidad de grandes volúmenes de datos clínicos y a las inversiones realizadas en innovación tecnológica. Sin embargo, diversos investigadores han señalado la necesidad de fortalecer aspectos relacionados con transparencia, protección de datos y supervisión ética (WHO, 2021).

En Colombia y otros países latinoamericanos, la adopción de inteligencia artificial en radiología se encuentra en una etapa de desarrollo más limitada. Aunque existen avances relacionados con transformación digital en salud, todavía persisten desafíos asociados a la regulación específica, la validación clínica local y la disponibilidad de infraestructura tecnológica. Esta situación genera la necesidad de fortalecer marcos regulatorios que permitan una implementación segura, ética y adaptada a las características de la población colombiana (INVIMA, 2023).

### **Estrategias de Mitigación de Sesgos y Riesgos Diagnósticos**

La identificación de sesgos algorítmicos y riesgos asociados al uso de inteligencia artificial ha impulsado el desarrollo de diversas estrategias orientadas a mejorar la seguridad y confiabilidad de estas herramientas. Una de las medidas más importantes consiste en garantizar

que los conjuntos de datos utilizados para entrenar los algoritmos sean amplios, diversos y representativos de la población a la cual estarán dirigidos (Dratsch et al., 2023).

Otra estrategia fundamental corresponde a la validación clínica continua de los sistemas antes y después de su implementación. Los algoritmos deben someterse a procesos rigurosos de evaluación que permitan identificar posibles diferencias de rendimiento entre distintos grupos poblacionales y contextos clínicos (Dratsch et al., 2023). La incorporación de modelos de inteligencia artificial explicable también representa una alternativa prometedora. Estos sistemas permiten comprender los criterios utilizados por el algoritmo para emitir determinadas recomendaciones, favoreciendo la transparencia y fortaleciendo la confianza de los profesionales de la salud (Chen et al., 2022). Desde el ámbito educativo, resulta indispensable promover programas de formación orientados al uso responsable de la inteligencia artificial. Los profesionales deben desarrollar competencias relacionadas con interpretación crítica de resultados, identificación de sesgos y comprensión básica del funcionamiento de los algoritmos. Esta preparación permitirá aprovechar los beneficios de la tecnología sin comprometer la calidad de la atención médica (Chen et al., 2022).

Finalmente, las instituciones sanitarias y los organismos reguladores deben establecer mecanismos permanentes de monitoreo, auditoría y actualización de los sistemas implementados. La inteligencia artificial es una tecnología dinámica que evoluciona constantemente, por lo que su supervisión debe mantenerse durante todo su ciclo de vida (Gichoya et al., 2023).

### **Síntesis del Marco Teórico**

La inteligencia artificial ha transformado progresivamente la práctica radiológica mediante herramientas capaces de optimizar el análisis de imágenes y apoyar la toma de decisiones clínicas. Tecnologías como el Machine Learning, el Deep Learning y los sistemas de

soporte a la decisión clínica han demostrado un importante potencial para mejorar la eficiencia diagnóstica y fortalecer los procesos asistenciales (Geis et al., 2019).

Sin embargo, la incorporación de estas herramientas también plantea desafíos significativos relacionados con la transferencia de sesgos, el sesgo de automatización y las implicaciones éticas derivadas de su utilización. La evidencia científica demuestra que los algoritmos pueden reproducir desigualdades presentes en los datos de entrenamiento y que los profesionales pueden verse influenciados por recomendaciones automatizadas cuando no mantienen una actitud crítica frente a los resultados generados por la tecnología (Goddard et al., 2012).

En este contexto, el papel del radiólogo continúa siendo fundamental. La inteligencia artificial debe entenderse como una herramienta de apoyo clínico que complementa el juicio profesional, pero no lo reemplaza. La integración responsable de estas tecnologías requiere procesos de formación continua, regulación adecuada, validación clínica permanente y estrategias orientadas a garantizar la equidad, la seguridad y la calidad de la atención médica (Topol, 2019).

Por lo tanto, comprender los riesgos asociados a la transferencia de sesgos y al sesgo de automatización resulta esencial para promover un uso ético y responsable de la inteligencia artificial en radiología, favoreciendo diagnósticos más precisos y una mejor atención para los pacientes (Geis et al., 2019).

## **Metodología**

El presente estudio se desarrolló bajo un enfoque cualitativo, debido a que busca comprender y analizar la información existente relacionada con la transferencia de sesgos y el sesgo de automatización en el uso de herramientas de inteligencia artificial aplicadas a la radiología. Este enfoque permite interpretar conceptos, perspectivas y hallazgos reportados en la literatura científica sobre el impacto de estas tecnologías en la práctica clínica y en la toma de decisiones diagnósticas (Hernández Sampieri & Mendoza, 2018).

La investigación corresponde a un estudio de tipo descriptivo y documental. Es descriptivo porque pretende identificar y caracterizar los principales aspectos relacionados con los sesgos presentes en los sistemas de inteligencia artificial utilizados en radiología, así como sus implicaciones clínicas, éticas y profesionales. Asimismo, es documental porque se fundamenta en la revisión y análisis de información proveniente de fuentes secundarias, tales como artículos científicos, publicaciones académicas, documentos institucionales y literatura especializada sobre inteligencia artificial en salud. El diseño metodológico es no experimental, ya que no se manipulan variables ni se realizan intervenciones directas sobre la población de estudio. La investigación se basa en la observación, recopilación y análisis de información previamente publicada, con el propósito de comprender el fenómeno estudiado desde una perspectiva teórica y crítica (Hernández Sampieri & Mendoza, 2018).

Como técnica principal de recolección de información se utilizó la revisión bibliográfica. Esta técnica permitió identificar, seleccionar y analizar documentos científicos relacionados con la inteligencia artificial, la radiología, los sesgos algorítmicos, el sesgo de automatización, la transferencia de sesgos y los aspectos éticos asociados a estas tecnologías. La información fue obtenida a partir de artículos indexados, revistas científicas, informes institucionales y

publicaciones de organismos internacionales reconocidos en el ámbito de la salud y la tecnología (World Health Organization, 2021). Las fuentes consultadas fueron seleccionadas teniendo en cuenta criterios de actualidad, relevancia científica, confiabilidad y relación directa con los objetivos de la investigación. Se priorizaron estudios publicados en los últimos años, debido al constante avance de la inteligencia artificial en el campo de la salud y particularmente en la radiología diagnóstica (European Society of Radiology, 2019).

### **Criterios de Inclusión y Exclusión**

Con el propósito de garantizar la pertinencia, calidad y actualidad de la información analizada, se establecieron criterios específicos para la selección de las fuentes documentales utilizadas en la presente investigación.

#### ***Criterios de Inclusión***

Se incluyeron artículos científicos publicados entre 2018 y 2025, con el fin de incorporar evidencia reciente relacionada con el desarrollo y aplicación de la inteligencia artificial en el ámbito de la salud. Asimismo, se seleccionaron estudios enfocados en la inteligencia artificial aplicada a la radiología, especialmente aquellos que abordaban el sesgo de automatización, la transferencia de sesgos, las implicaciones éticas y la toma de decisiones clínicas. También se consideraron publicaciones en español e inglés debido a su relevancia en la producción científica internacional sobre el tema. Finalmente, se incluyeron únicamente artículos revisados por pares y documentos provenientes de organismos e instituciones reconocidas por su rigor académico y científico.

#### ***Criterios de Exclusión***

Se excluyeron documentos duplicados identificados durante el proceso de búsqueda bibliográfica. Asimismo, se descartó la literatura que no guardaba relación directa con la radiología

o con el uso de inteligencia artificial en contextos clínicos. También fueron excluidos los artículos sin acceso al texto completo, debido a la imposibilidad de realizar una evaluación adecuada de su contenido. Finalmente, se eliminaron fuentes no científicas, documentos de opinión sin respaldo académico y publicaciones que no cumplieran criterios mínimos de calidad metodológica.

## **Fases de la Investigación**

### ***Fase 1. Búsqueda y Recopilación de la Información***

En esta fase se realizó la identificación y búsqueda de literatura científica relacionada con la inteligencia artificial aplicada a la radiología, los sesgos algorítmicos, el sesgo de automatización y la transferencia de sesgos. Para ello se consultaron artículos científicos, revistas especializadas, documentos institucionales y publicaciones académicas disponibles en bases de datos reconocidas y fuentes confiables (Pesapane et al., 2018).

### ***Fase 2. Selección y Organización de la Información***

Posteriormente, se efectuó una revisión detallada de los documentos recopilados con el fin de seleccionar aquellos que presentaban mayor pertinencia para el desarrollo del estudio. La información fue organizada de acuerdo con las categorías temáticas establecidas en los objetivos de investigación, permitiendo estructurar de manera coherente los contenidos relacionados con los sesgos en inteligencia artificial y su impacto en radiología (Geis et al., 2019).

### ***Fase 3. Análisis e Interpretación de la Información***

Una vez organizada la información, se procedió al análisis crítico de los contenidos encontrados. En esta etapa se identificaron conceptos, hallazgos, problemáticas y aportes relevantes sobre la influencia de la transferencia de sesgos y el sesgo de automatización en la toma de decisiones clínicas apoyadas por sistemas de inteligencia artificial. Este proceso

permitió establecer relaciones entre los diferentes planteamientos de los autores consultados y responder a la pregunta de investigación propuesta (Cabitza et al., 2017).

#### ***Fase 4. Elaboración del Documento Final***

Finalmente, se realizó la redacción del informe de investigación, integrando los resultados obtenidos a partir de la revisión documental y el análisis de la literatura científica. Esta fase permitió consolidar los hallazgos encontrados y formular conclusiones relacionadas con el uso responsable, ético y seguro de la inteligencia artificial en el campo de la radiología (World Health Organization, 2021).

## Resultados

### **El Sesgo de Automatización: Cuando el Algoritmo Reemplaza al Juicio Clínico**

Para comprender a fondo el sesgo de automatización, es necesario partir de su definición conceptual. Este tipo de sesgo describe la inclinación de las personas a sobrevalorar las recomendaciones producidas por sistemas automatizados, incluso cuando estas son incorrectas o cuestionables. En el ámbito radiológico, esta situación ocurre cuando el radiólogo adopta la categoría diagnóstica sugerida por una herramienta de IA sin realizar una verificación independiente y suficiente de la imagen.

La evidencia empírica sobre este problema es contundente. En un estudio prospectivo publicado en la revista *Radiology*, Dratsch et al. (2023) evaluaron a 27 radiólogos de distintos niveles de experiencia mientras interpretaban 50 mamografías con el apoyo de un sistema simulado de IA. Los resultados mostraron que, en los casos en que la herramienta sugería una categoría BI-RADS incorrecta, la precisión diagnóstica de los radiólogos sin experiencia descendía desde aproximadamente el 80% hasta menos del 20%. Incluso los lectores más experimentados con más de quince años de práctica vieron caer su exactitud desde el 82% hasta el 45.5% bajo la influencia de sugerencias erróneas del sistema. Este hallazgo resulta especialmente relevante porque demuestra que el sesgo de automatización no es exclusivo de los profesionales con poca formación, sino que afecta también a quienes cuentan con amplia trayectoria clínica. La lógica detrás de este fenómeno es comprensible: la IA se percibe como una herramienta objetiva y altamente procesadora de información, lo que genera en el profesional una confianza desproporcionada que puede inhibir la aplicación del razonamiento crítico independiente.

Desde una perspectiva cognitiva, investigaciones como la de Lyell y Coiera (2017), publicada en el *Journal of the American Medical Informatics Association*, identificaron mediante una revisión sistemática que la complejidad del proceso de verificación es uno de los factores que más influye en la aparición del sesgo de automatización. Cuando el profesional debe corroborar manualmente una sugerencia automatizada, el esfuerzo cognitivo adicional puede llevar a omitir ese paso, especialmente en entornos de alta carga de trabajo como los servicios de urgencias radiológicas.

Otro matiz importante que arroja la literatura es el papel de la actitud previa del profesional hacia la IA. Un estudio de Rezazade Mehrizi et al. (2023), publicado en *Scientific Reports*, analizó 2.760 decisiones de 92 radiólogos y encontró que tanto la actitud positiva como la negativa hacia la IA puede modelar la forma en que el profesional interactúa con las sugerencias del algoritmo. Los radiólogos que poseían una actitud favorable tendían a seguir las recomendaciones del sistema incluso cuando estas eran equivocadas, mientras que aquellos con una actitud más crítica mostraban mayor independencia diagnóstica, aunque también mayor propensión a rechazar sugerencias correctas.

### **La Transferencia de Sesgo: Cuando los Datos Reproducen Inequidades**

La transferencia de sesgo en radiología con IA representa un problema de naturaleza estructural que va más allá del comportamiento individual del profesional. Este fenómeno se origina en las etapas de recolección y curación de los datos de entrenamiento, cuando los conjuntos de imágenes utilizados para desarrollar los modelos no reflejan de manera proporcional la diversidad de la población que los sistemas deberán atender en el mundo real.

Banerjee et al. (2023), en su investigación sobre sesgo demográfico en modelos de lenguaje visual aplicados a imágenes médicas, documentaron que modelos de IA de nivel

experto consistentemente subdiagnosticaba a grupos históricamente marginados, como mujeres de raza negra, en comparación con radiólogos certificados. Este hallazgo fue replicado en cinco conjuntos de datos globales de radiografía de tórax, lo que refuerza la hipótesis de que se trata de un patrón sistemático y no de un resultado aislado. Por ejemplo, algunos sistemas de inteligencia artificial entrenados principalmente con imágenes de poblaciones europeas o norteamericanas pueden presentar menor precisión diagnóstica en pacientes latinoamericanos o afrodescendientes. Esto podría generar subestimación de patologías pulmonares, retrasos diagnósticos o errores en la clasificación de enfermedades, afectando especialmente a grupos históricamente subrepresentados en los conjuntos de datos médicos. En esta misma línea, Gichoya et al. (2023), en su revisión orientada hacia la equidad en la IA para imágenes médicas, señalaron que los sesgos pueden introducirse en múltiples etapas del ciclo de vida de un modelo: durante la selección de los datos, en el proceso de etiquetado, en el diseño de la arquitectura del modelo y, finalmente, en su despliegue clínico.

Esta cadena de transmisión de sesgo dificulta su identificación y en esta misma línea, Gichoya et al. (2023), en su revisión orientada hacia la equidad en la IA para imágenes médicas, señalaron que los sesgos pueden introducirse en múltiples etapas del ciclo de vida de un modelo: durante la selección de los datos, en el proceso de etiquetado, en el diseño de la arquitectura del modelo y, finalmente, en su despliegue clínico. Esta cadena de transmisión de sesgo dificulta su identificación y corrección, porque en muchos casos los desarrolladores y los usuarios finales no tienen visibilidad sobre todos los pasos del proceso. Un aspecto que frecuentemente se pasa por alto es la influencia de las decisiones técnicas durante el desarrollo del modelo. Chen et al. (2022), en un artículo publicado en *Radiology: Artificial Intelligence*, analizaron el impacto del aprendizaje por transferencia, una técnica ampliamente utilizada en radiología para entrenar

modelos con conjuntos de datos de tamaño limitado y advirtieron que esta estrategia puede transferir también los sesgos presentes en los modelos pre-entrenados utilizados como punto de partida. Esto significa que incluso cuando se dispone de datos locales relativamente representativos, el modelo heredado de otra fuente puede contaminar el nuevo sistema con sus propias distorsiones.

La dimensión geográfica del problema también merece atención. Estudios como el revisado por Larson et al. evidencian que la mayoría de los conjuntos de datos de referencia en imagenología médica provienen de instituciones de Estados Unidos o China, lo que implica que los modelos entrenados con esos datos pueden tener un rendimiento notablemente inferior cuando se aplican en poblaciones latinoamericanas, africanas o del sudeste asiático. Para países como Colombia, donde las condiciones epidemiológicas, los equipos utilizados y las características demográficas difieren significativamente de las de los contextos de origen, esta transferencia de sesgo representa un riesgo diagnóstico concreto.

En consecuencia, la calidad y representatividad de los datos utilizados para entrenar los algoritmos tiene un impacto directo en la práctica clínica, ya que los resultados emitidos por la inteligencia artificial pueden influir en las decisiones diagnósticas y terapéuticas tomadas por el profesional de la salud.

### **Implicaciones Éticas: Autonomía del Paciente y Justicia Algorítmica**

El análisis de los sesgos en inteligencia artificial aplicada a radiología no puede limitarse únicamente a aspectos técnicos o clínicos, sino que también requiere una reflexión ética sobre los principios que orientan la atención en salud. Entre ellos destacan la autonomía del paciente, la justicia distributiva y la no maleficencia (Vokinger et al., 2021). En relación con la autonomía, el uso de sistemas de IA plantea interrogantes sobre el derecho del paciente a conocer cuándo una

herramienta automatizada participa en el proceso diagnóstico. La literatura señala que la falta de transparencia puede afectar la toma de decisiones informadas y limitar la comprensión de los posibles riesgos asociados al uso de algoritmos clínicos (Gichoya et al., 2023). Respecto a la justicia algorítmica, diversos estudios han evidenciado que algunos sistemas de IA presentan diferencias de rendimiento entre grupos poblacionales, especialmente cuando fueron entrenados con bases de datos poco representativas. Esta situación puede generar desigualdades diagnósticas que afecten a poblaciones históricamente subrepresentadas, lo que constituye un desafío ético para los sistemas de salud (Banerjee et al., 2023; Vokinger et al., 2021).

Finalmente, el principio de no maleficencia exige que las instituciones sanitarias implementen mecanismos de supervisión y evaluación continua de estas herramientas, garantizando que su utilización no genere daños derivados de errores sistemáticos o sesgos algorítmicos (WHO, 2021).

### **Colaboración Interdisciplinaria: Un Requisito Ineludible**

La implementación responsable de la inteligencia artificial en radiología requiere la participación conjunta de profesionales de diferentes disciplinas. La colaboración entre radiólogos, científicos de datos, ingenieros, expertos en bioética y responsables de políticas sanitarias permite identificar riesgos potenciales y desarrollar estrategias para reducir los sesgos presentes en los sistemas automatizados (Gichoya et al., 2023). Los radiólogos aportan conocimientos clínicos fundamentales para la interpretación de imágenes y la validación de resultados, mientras que los científicos de datos contribuyen al desarrollo y optimización de algoritmos. La integración de ambas perspectivas favorece la creación de herramientas más seguras y adaptadas a las necesidades reales de la práctica médica (Chen et al., 2022). Asimismo, la participación de expertos en bioética permite evaluar las implicaciones sociales y morales

derivadas del uso de la IA en salud. La literatura destaca que la inclusión de pacientes y representantes comunitarios en los procesos de evaluación contribuye a fortalecer la transparencia y la legitimidad de estas tecnologías (WHO, 2021).

### **Desafíos Prácticos: Obstáculos Regulatorios, Económicos y de Carga Laboral**

La incorporación de sistemas de inteligencia artificial en radiología enfrenta diversos desafíos relacionados con aspectos económicos, regulatorios y organizacionales. Uno de los principales obstáculos es el alto costo asociado a la recopilación, almacenamiento y etiquetado de grandes volúmenes de imágenes médicas necesarias para el entrenamiento de algoritmos confiables (Chen et al., 2022). Por otra parte, la elevada carga laboral de muchos servicios de radiología puede favorecer la aparición del sesgo de automatización, ya que los profesionales podrían depender excesivamente de las recomendaciones generadas por los sistemas automatizados para agilizar los procesos diagnósticos (Lyell & Coiera, 2017). A nivel regulatorio, diversos autores señalan que aún existen vacíos normativos relacionados con la evaluación de equidad, transparencia y seguridad de las herramientas de IA utilizadas en medicina. Por ello, se considera necesario fortalecer los marcos regulatorios para garantizar una implementación responsable de estas tecnologías (FDA, 2024; European Commission, 2024).

### **Evolución del Rol Profesional del Radiólogo**

La creciente integración de la inteligencia artificial en los servicios de imagenología está transformando el papel tradicional del radiólogo. Lejos de reemplazar al profesional, la evidencia científica sugiere que estas tecnologías funcionan principalmente como herramientas de apoyo para mejorar la precisión y eficiencia diagnóstica (Rezazade Mehrizi et al., 2023). En este contexto, el radiólogo adquiere nuevas responsabilidades relacionadas con la supervisión crítica de los resultados generados por la IA, la identificación de posibles errores algorítmicos y la

integración de la información clínica del paciente en la toma de decisiones diagnósticas (Dratsch et al., 2023). Además, diversos autores destacan la necesidad de fortalecer la formación en competencias digitales y alfabetización en inteligencia artificial dentro de los programas de educación médica, con el fin de preparar a los profesionales para interactuar adecuadamente con estas herramientas emergentes (WHO, 2021). Paralelamente, la IA abre posibilidades para que el radiólogo intensifique su rol en la relación directa con el paciente y con el equipo clínico. En la medida en que las tareas más rutinarias quedan parcialmente automatizadas, el profesional puede dedicar más tiempo a la comunicación de hallazgos complejos, a la participación en reuniones de comité de tumores o a la asesoría interdisciplinaria. Esto implica una revalorización de habilidades comunicativas y relacionales que el modelo clásico de la radiología como especialidad "de sala oscura" tendía a relegar a un segundo plano (Beam & Kohane, 2018).

No obstante, esta transformación solo será positiva si va acompañada de procesos formativos adecuados y de condiciones institucionales que la hagan viable. Un radiólogo que interactúa cotidianamente con herramientas de IA sin comprender sus limitaciones, sus posibles sesgos y los principios básicos de su funcionamiento es un profesional vulnerable al sesgo de automatización. La alfabetización en IA no es, por tanto, una competencia opcional para las nuevas generaciones de especialistas en imagenología: es una condición de su práctica responsable (OPS, 2023).

## **Marcos Regulatorios: Un Análisis Comparativo Internacional y el Contexto**

### **Latinoamericano**

La pregunta sobre cómo regular la inteligencia artificial (IA) en salud ha generado respuestas diversas en distintas regiones del mundo. La comparación entre estos enfoques permite identificar tanto las mejores prácticas disponibles como las brechas regulatorias que

persisten, particularmente en América Latina (WHO, 2021). En los Estados Unidos, la Administración de Alimentos y Medicamentos (FDA) ha sido uno de los primeros organismos en desarrollar lineamientos específicos para la regulación de software basado en inteligencia artificial como dispositivo médico. A través de su programa para sistemas de aprendizaje automático y aprendizaje profundo, la FDA reconoce que estos modelos pueden modificar su comportamiento con el tiempo y propone una supervisión continua durante todo el ciclo de vida del producto (FDA, 2024). Sin embargo, diversos autores señalan que los requisitos relacionados con la evaluación de equidad y sesgo algorítmico aún son limitados, ya que no siempre se exige la presentación de resultados diferenciados según características demográficas de la población usuaria (Vokinger et al., 2021).

En la Unión Europea, el Reglamento de Inteligencia Artificial (AI Act), aprobado en 2024, clasifica las aplicaciones médicas de IA como sistemas de alto riesgo, lo que implica requisitos estrictos de transparencia, trazabilidad, supervisión humana y gestión de riesgos (European Commission, 2024). Este marco regulatorio incorpora además criterios relacionados con la prevención de la discriminación y la protección de los derechos fundamentales, convirtiéndose en una de las normativas más avanzadas a nivel internacional (European Commission, 2024).

Por su parte, el Reino Unido ha impulsado mecanismos de evaluación a través del National Health Service (NHS) y del National Institute for Health and Care Excellence (NICE). Estas instituciones han desarrollado estándares que buscan garantizar la seguridad, eficacia y equidad de las herramientas de inteligencia artificial utilizadas en entornos clínicos, exigiendo evidencia científica que respalde su desempeño en diferentes grupos poblacionales (NICE, 2022). En contraste, América Latina presenta un desarrollo regulatorio más limitado. En

Colombia, aunque existen normativas relacionadas con dispositivos médicos y software sanitario, todavía no se cuenta con una regulación específica para la evaluación de sesgos algorítmicos en sistemas de inteligencia artificial aplicados a radiología (INVIMA, 2023). De manera similar, otros países de la región continúan construyendo marcos regulatorios adaptados a los desafíos que plantea la transformación digital en salud. Brasil constituye uno de los casos más avanzados de la región. La Agencia Nacional de Vigilancia Sanitaria (ANVISA) ha desarrollado lineamientos para software médico que incluyen aspectos relacionados con trazabilidad, seguridad y gestión de riesgos tecnológicos. No obstante, la literatura especializada señala que aún existen desafíos asociados a la evaluación sistemática de sesgos y a la validación clínica en poblaciones diversas (WHO, 2021).

La situación latinoamericana resulta especialmente relevante debido a que gran parte de las herramientas de inteligencia artificial utilizadas en la región han sido desarrolladas y entrenadas en países del Norte Global. Esto incrementa el riesgo de transferencia de sesgos, ya que los algoritmos pueden no reflejar adecuadamente las características demográficas, epidemiológicas y socioculturales de las poblaciones latinoamericanas (Gichoya et al., 2023). Por esta razón, diversos autores destacan la necesidad de implementar procesos de validación local antes de la adopción clínica de estas tecnologías (Banerjee et al., 2023).

### **Ejercicio de Perspectivas Internacionales**

La implementación de la inteligencia artificial en radiología ha generado transformaciones significativas en distintos sistemas de salud alrededor del mundo. Países como Estados Unidos, Reino Unido y China lideran actualmente el desarrollo de herramientas orientadas a la detección temprana de enfermedades, el análisis automatizado de imágenes médicas y la optimización de procesos diagnósticos (Chen et al., 2022). En Estados Unidos,

numerosas instituciones académicas y hospitalarias han impulsado investigaciones centradas en el uso de algoritmos de aprendizaje profundo para mejorar la precisión diagnóstica en mamografía, tomografía computarizada y radiografía convencional (Dratsch et al., 2023). Paralelamente, la FDA ha fortalecido los mecanismos regulatorios destinados a supervisar estas tecnologías (FDA, 2024). En Europa, la entrada en vigor del AI Act representa un avance importante en la construcción de un modelo regulatorio enfocado en la transparencia, la supervisión humana y la reducción de riesgos éticos asociados al uso de sistemas automatizados (European Commission, 2024). Por otra parte, China ha experimentado un rápido crecimiento en el desarrollo de soluciones basadas en inteligencia artificial gracias a la disponibilidad de grandes volúmenes de datos clínicos y al respaldo institucional para la investigación tecnológica. Sin embargo, diversos estudios advierten que la expansión acelerada de estas herramientas también puede aumentar problemas relacionados con sesgos algorítmicos y desigualdades diagnósticas si no se implementan mecanismos adecuados de control y validación (Gichoya et al., 2023).

Desde una perspectiva internacional, la evidencia muestra que la inteligencia artificial tiene el potencial de fortalecer la calidad diagnóstica y optimizar los servicios de radiología. No obstante, su implementación requiere marcos éticos, regulatorios y científicos sólidos que permitan minimizar riesgos asociados al sesgo de automatización y a la transferencia de sesgos en los sistemas de salud (WHO, 2021; Vokinger et al., 2021).

### **Implicaciones Clínicas y Estrategias de Mitigación**

Reconocer la existencia del sesgo de automatización y de la transferencia de sesgo no es suficiente si no se desarrollan estrategias concretas para reducir su impacto en la práctica clínica.

Diversos estudios proponen intervenciones dirigidas tanto al diseño tecnológico como a la formación de los profesionales y la organización de los servicios de salud (Dratsch et al., 2023).

Una de las estrategias más recomendadas consiste en mostrar niveles de confianza asociados a cada recomendación emitida por el algoritmo. Esta información permite que el radiólogo valore de manera más crítica los resultados y disminuya el riesgo de dependencia excesiva de la herramienta automatizada (Dratsch et al., 2023). Asimismo, el desarrollo de sistemas de inteligencia artificial explicable (Explainable Artificial Intelligence, XAI) favorece una mayor comprensión de los criterios utilizados por el algoritmo para emitir una conclusión diagnóstica. La transparencia del proceso contribuye a fortalecer la supervisión humana y a reducir el sesgo de automatización (Chen et al., 2022).

En cuanto a la transferencia de sesgo, las estrategias más utilizadas incluyen la diversificación de los conjuntos de datos, la incorporación de muestras representativas de diferentes grupos poblacionales y el monitoreo continuo del rendimiento de los modelos después de su implementación clínica (Gichoya et al., 2023). Stanley et al. (2024) demostraron que la reponderación de grupos subrepresentados constituye una de las alternativas más prometedoras para disminuir diferencias de rendimiento entre poblaciones. Finalmente, la formación de los profesionales de la salud continúa siendo uno de los pilares fundamentales para la mitigación de riesgos. La Organización Mundial de la Salud (WHO, 2021) recomienda incorporar contenidos relacionados con inteligencia artificial, ética digital y evaluación crítica de algoritmos dentro de los programas de educación médica y formación continua, con el propósito de garantizar un uso responsable y seguro de estas tecnologías.

## Conclusiones

La incorporación de sistemas basados en IA ha supuesto una transformación profunda y positiva en la práctica radiológica al ofrecer capacidades sin precedentes para procesar grandes volúmenes de información, identificar patrones complejos en imágenes médicas, optimizando tiempos de interpretación y apoyando la detección temprana de patologías. No obstante, la evidencia revisada confirma de manera consistente que estas tecnologías deben concebirse y utilizarse exclusivamente como instrumentos de apoyo clínico, diseñados para amplificar, fortalecer y agilizar el trabajo del profesional, pero nunca para reemplazarlo. El razonamiento clínico, la integración de datos del historial del paciente, la comprensión del contexto epidemiológico y la toma de decisiones responsables siguen siendo competencias irrenunciables del radiólogo fundamentadas en su formación, experiencia y criterio ético. La IA aporta eficiencia y precisión técnica, pero carece de la capacidad humana para interpretar la complejidad integral de cada caso y para asumir la responsabilidad última sobre el diagnóstico y su impacto en la vida del paciente.

La transferencia de sesgos se identifica como uno de los desafíos más críticos y estructurales del uso de la IA en radiología, con implicaciones que van más allá de lo técnico para afectar principios fundamentales de justicia y equidad en salud. Este fenómeno se origina principalmente en la calidad, representación y procedencia de los datos utilizados para entrenar los algoritmos, cuando las bases de datos se construyen con información limitada, sesgada o concentrada en poblaciones específicas generalmente de regiones desarrolladas con características demográficas, genéticas y epidemiológicas distintas a las de América Latina, los sistemas aprenden y replican estas limitaciones, generando resultados menos precisos o incluso erróneos al aplicarse a poblaciones diversas como la nuestra.

El uso de IA en radiología plantea retos éticos que son inseparables de su desarrollo y aplicación. Se concluye que la transparencia es un requisito indispensable tanto en profesionales como pacientes. Estos tienen derecho a comprender cuándo y cómo interviene un sistema automatizado en su diagnóstico, así como a conocer sus limitaciones. La autonomía del paciente exige que se respete su derecho a recibir información clara y a participar en las decisiones sobre su atención. Asimismo, se reafirma que la responsabilidad por el diagnóstico y sus consecuencias recae siempre en el profesional de la salud, sin que la existencia de una recomendación automatizada pueda transferir o diluir esta obligación. Además, la ética sanitaria exige que se garanticen condiciones de equidad, evitando que la tecnología se convierta en una fuente de discriminación o de acceso desigual a servicios de calidad. En este sentido, la IA en radiología debe evaluarse no solo por su eficacia técnica, sino también por su capacidad para respetar y promover los valores humanos y los principios de la atención sanitaria.

El análisis comparativo internacional evidencia que, en países como Estados Unidos, los integrantes de la Unión Europea o el Reino Unido han avanzado en marcos regulatorios específicos, orientados a garantizar la seguridad, equidad y transparencia de estas herramientas. En contraste, en Colombia y en gran parte de América Latina, la regulación se encuentra en una etapa inicial, con vacíos normativos importantes en lo que respecta a la evaluación de sesgos, la validación en poblaciones locales y la supervisión continua de los sistemas de IA. Se concluye que esta situación representa un riesgo adicional, ya que las herramientas que se implementan en la región suelen haber sido diseñadas y validadas en contextos diferentes, sin adaptarse a nuestras características epidemiológicas, demográficas y sanitarias.

### Referencias Bibliográficas

- Banerjee, I., et al. (2023). *Reading race: AI recognises patient race in medical images*. The Lancet Digital Health, 5(2), e71–e82. [https://doi.org/10.1016/S2589-7500\(22\)00230-8](https://doi.org/10.1016/S2589-7500(22)00230-8)
- Beam, A. L., & Kohane, I. S. (2018). *Big data and machine learning in health care*
- Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). *Unintended consequences of machine learning in medicine*.
- Chen, I. Y., Joshi, S., Ghassemi, M., & others. (2022). *Treating health disparities with artificial intelligence*. Nature Medicine, 26(1), 16–17. <https://doi.org/10.1038/s41591-019-0649-2>
- Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic medicine*, 78(8), 775-780.
- D., Shahidi, R., Fatehi, M., & Dratsch, T. (2023). *The impact of AI suggestions on radiologists' decisions: A pilot study of explainability and attitudinal priming interventions in mammography examination*. Scientific Reports, 13, 9230. <https://doi.org/10.1038/s41598-023-36435-3>
- Davenport, T., & Kalakota, R. (2019). *The potential for artificial intelligence in healthcare*.
- Dratsch, T., Chen, X., Rezazade Mehrizi, M., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D., & Pinto Dos Santos, D. (2023). *Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance*. Radiology, 307(4), e222176. <https://doi.org/10.1148/radiol.222176>
- Dratsch, T., Chen, X., Rezazade Mehrizi, M., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D., & Pinto Dos Santos, D. (2023). *Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance*. Radiology, 307(2), e222176. <https://doi.org/10.1148/radiol.222176>.

- Esteva, A., et al. (2017). *Dermatologist-level classification of skin cancer with deep neural networks*. *Nature*.
- European Parliament. (2024). Artificial Intelligence Act (AI Act). European Union
- Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., ... & Kohli, M. D. (2019). *Ethics of artificial intelligence in radiology: Summary of the Joint European and North American Multisociety Statement*. *Radiology*, 293(2), 436–440.
- Gichoya, J. W., Banerjee, I., Bhimireddy, A. R., Burns, J. L., Celi, L. A., Chen, L. C., Correa, R., Dullerud, N., Ghassemi, M., Huang, S. C., & others. (2023). *AI recognition of patient race in medical imaging: A modelling study*. *The Lancet Digital Health*, 4(6), e406–e414. [https://doi.org/10.1016/S2589-7500\(22\)00063-2](https://doi.org/10.1016/S2589-7500(22)00063-2)
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). *Automation bias: A systematic review of frequency, effect mediators, and mitigators*. <https://arxiv.org/abs/1711.05225>.
- Green, D., Swets, M. (1966). *Signal detection theory and psychophysics*.
- Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide*. Routledge.
- INVIMA. (2023). *Lineamientos para software como dispositivo médico en Colombia*. Instituto Nacional de Vigilancia de Medicamentos y Alimentos
- Jha, S., & Topol, E. J. (2016). *Adapting to artificial intelligence: Radiologists and pathologists as information specialists*. *JAMA*, 316(22), 2353–2354.
- Larson, D. B., Harvey, H., Rubin, D. L., Irani, N., Tse, J. R., & Langlotz, C. P. (2021). *Regulatory frameworks for development and evaluation of artificial intelligence–based diagnostic imaging algorithms: Summary and recommendations*. *Journal of the American College of Radiology*, 18(3), 413–424. <https://doi.org/10.1016/j.jacr.2020.09.060>

- Litjens, G., et al. (2017). *A survey on deep learning in medical image analysis*.
- Lyell, D., & Coiera, E. (2017). *Automation bias and verification complexity: A systematic review*. *Journal of the American Medical Informatics Association*, 24(2), 423
- Menchú López, W. E. (2025). *Inteligencia artificial en radiología*. *Revista Médica Digital*, 14(2), 45–58.
- Mora Chacón, S. (2024). *imágenes médicas*. Editorial Académica de Salud.
- Morley, J., Machado, C. C. V., Burr, C., Cows, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). *The ethics of AI in health care: A mapping review*.
- Nair, A. V., et al. (2022). *Artificial intelligence in radiology: Current status and future directions*. *Radiology Journal*, 15(3), 120–135.
- National Institute for Health and Care Excellence (NICE). (2022). *Evidence standards framework for digital health technologies*. NICE.
- Beam, A. L., & Kohane, I. S. (2018). *Big data and machine learning in health care*. *JAMA*, 319(13), 1317–1318. <https://doi.org/10.1001/jama.2017.18391>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). *Dissecting racial bias in an algorithm used to manage the health of populations*. *Science*, 366(6464), 447–453.
- Organización Panamericana de la Salud (OPS). (2023). *Inteligencia artificial y transformación digital en salud en las Américas*. OPS.
- Parikh, R. B., Teeple, S., & Navathe, A. S. (2019). *Addressing bias in artificial intelligence in health care*. *JAMA*, 322(24), 2377–2378.
- Pesapane, F., Codari, M., & Sardanelli, F. (2018). *Artificial intelligence in medical imaging*. *European Radiology Experimental*, 2(1), 35.
- Rajpurkar, P., et al. (2017). *CheXNet: Radiologist-level pneumonia detection*.

Rezazade Mehrizi, M. H., Mol, F., Peter, M., Ranschaert, E., Pinto Dos Santos.

Stanley, E., et al. (2024). *Frameworks for objective bias assessment in medical imaging artificial intelligence*. *Radiology: Artificial Intelligence*, 6(1), e230145.

<https://doi.org/10.1148/ryai.230145>

Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. *Basic Books*.

Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. (2021). *Mitigating bias in machine learning for medicine*. *Communications Medicine*, 1(25), 1–3.

<https://doi.org/10.1038/s43856-021-00028-w>.

