# Evaluación de la integridad diagnostica frente a la generación de hallazgos patológicos falsos mediante IA: revisión documental de la literatura actual

Ana Lorena Hurtado García

Denix Zulay Albarracín Arias

Lina María Aristizábal García

Jorge Orlando Lizarazo Marín

Juan Andrés Ruíz Blanco

Asesor

Edna Rocío Jamaica Guío

Universidad Nacional Abierta y a Distancia - UNAD

Escuela de Ciencias de la Salud - ECISA

Tecnología en Radiología e Imágenes Diagnósticas

2026

## Dedicatoria

Dedicamos este logro a nuestras familias, que con su apoyo incondicional, comprensión y motivación constante nos acompañaron en cada etapa de este proceso. Su confianza fue fundamental para avanzar con determinación en el desarrollo de este diplomado.

A nuestros tutores, por su compromiso para fomentar la responsabilidad en nosotros, lo que nos hace mejores personas y profesionales.

A todas las personas que, de una u otra manera, contribuyeron con su orientación, palabra de aliento o colaboración.

Por último, a nosotros, por la disciplina, el esfuerzo y la perseverancia para culminar satisfactoriamente este desafío profesional: Ana Lorena Hurtado García, Denix Zulay Albarracín Arias, Lina María Aristizábal García, Jorge Orlando Lizarazo Marín y Juan Andrés Ruíz Blanco.

**Agradecimientos**

Los autores expresamos nuestro sincero agradecimiento a Dios, por brindarnos la fortaleza, la sabiduría y la resiliencia necesarias para asumir cada reto presentado durante el desarrollo de este diplomado.

A nuestras familias, por su apoyo permanente, su paciencia inagotable y su confianza, fueron nuestro soporte e impulso para el crecimiento académico y personal.

A la UNAD, por brindarnos los espacios y recursos para desarrollar este proceso de formación, con una educación flexible, ética, autónoma y de calidad.

Expresamos nuestra gratitud a los tutores, principalmente a Edna Rocío Jamaica Guio y Christian Camilo Rodríguez Castro, que con su conocimiento y acompañamiento continuo contribuyeron al fortalecimiento de nuestras competencias profesionales.

Finalmente, agradecemos a cada uno de los integrantes de este equipo de trabajo, por su compromiso, responsabilidad y disposición para alcanzar juntos este objetivo.

## Resumen

La integración de IA y la radiología digital ha optimizado los procesos diagnósticos, pero también ha incrementado los riesgos de ciberseguridad, especialmente frente a ataques adversariales de redes generativas antagónicas (GAN), que modifican imágenes médicas e insertan hallazgos patológicos falsos, comprometiendo la autenticidad de los estudios, la integridad del registro clínico y la seguridad del paciente. El estudio es de revisión documental con enfoque cualitativo y diseño no experimental. Se analizaron artículos científicos, reportes técnicos, normativas internacionales y literatura en ciberseguridad hospitalaria, IA médica y sistemas de gestión de imágenes. La revisión identificó vulnerabilidades algorítmicas, fallas en la infraestructura tecnológica y riesgos asociados a redes hospitalarias interconectadas, además de brechas regulatorias y éticas respecto a la responsabilidad profesional y la protección de datos. Los hallazgos revelan debilidades que facilitan la inserción de artefactos sintéticos capaces de alterar diagnósticos y afectar los procesos clínicos. Asimismo, los sistemas PACS y redes hospitalarias amplían la superficie de ataque, permitiendo manipulaciones que afectan la trazabilidad, confiabilidad institucional y calidad del servicio radiológico. Los ataques adversariales y los deepfakes médicos son una amenaza creciente, con implicaciones legales, administrativas y operativas. Se concluye que se deben fortalecer los marcos de ciberseguridad, robustecer los algoritmos, implementar mecanismos de verificación de autenticidad y establecer regulaciones que garanticen la integridad de la información y la seguridad del paciente. La manipulación adversarial mediante GAN es una amenaza que exige estrategias integrales de protección, gobernanza tecnológica y mejora continua.


***Palabras clave:*** ciberseguridad, IA, ataques adversariales, redes generativas antagónicas (GAN), deepfakes médicos.

## Abstract

The integration of AI and digital radiology has optimized diagnostic processes but has also increased cybersecurity risks, particularly in the face of adversarial attacks generated through Generative Adversarial Networks (GANs). These attacks modify medical images and insert false pathological findings, compromising study authenticity, clinical record integrity, and patient safety. This study is a documentary review with a qualitative, non-experimental design. Scientific articles, technical reports, international regulations, and literature on hospital cybersecurity, medical AI, and imaging management systems were analyzed. The review identified algorithmic vulnerabilities, technological infrastructure failures, and risks associated with interconnected hospital networks, as well as regulatory and ethical gaps related to professional responsibility and data protection. The findings reveal weaknesses that facilitate the insertion of synthetic artifacts capable of altering diagnoses and affecting clinical processes. Likewise, PACS systems and hospital networks expand the attack surface, enabling manipulations that undermine traceability, institutional reliability, and the quality of radiological services. Adversarial attacks and medical deepfakes represent a growing threat with legal, administrative, and operational implications. The study concludes that cybersecurity frameworks must be strengthened, algorithms reinforced, authenticity-verification mechanisms implemented, and regulations established to ensure information integrity and patient safety. Adversarial manipulation through GANs constitutes a threat that demands comprehensive protection strategies, technological governance, and continuous improvement.


***Keywords:*** cybersecurity, AI, adversarial attacks, generative adversarial networks (GAN), medical deepfakes.

**Tabla de Contenido**

## Lista de Tablas

## Lista de Figuras

## Introducción

El avance de la inteligencia artificial (IA) en la radiología digital ha transformado significativamente los procesos de análisis, interpretación y gestión de imágenes médicas, permitiendo mejorar la precisión diagnóstica, optimizar los tiempos de lectura y fortalecer la eficiencia clínica.

Sin embargo, esta evolución tecnológica también ha incrementado la superficie de ataque en los sistemas de salud, dando lugar a nuevas amenazas asociadas a la ciberseguridad, especialmente mediante técnicas de manipulación adversarial como las redes generativas antagónicas (GAN), los ataques adversariales y el envenenamiento de datos, que tienen la capacidad de generar imágenes médicas sintéticas altamente realistas o modificar estudios existentes mediante la inserción, alteración o eliminación de estructuras patológicas de forma imperceptible al ojo humano. Estas modificaciones, conocidas como deepfakes médicos o ataques adversariales, representan un riesgo significativo para la integridad del registro clínico, ya que pueden inducir diagnósticos erróneos, alterar la evidencia médica y comprometer la toma de decisiones clínicas. La gravedad de esta problemática radica en que dichas alteraciones pueden evadir los mecanismos tradicionales de detección, afectando directamente la confiabilidad de los sistemas de apoyo diagnóstico basados en IA.

Estas tecnologías tienen la capacidad de generar imágenes médicas sintéticas altamente realistas o modificar estudios existentes de manera imperceptible, insertando, alterando o eliminando hallazgos patológicos sin dejar evidencia visible. Esta situación representa un riesgo crítico para la integridad del registro clínico, la confiabilidad diagnóstica y la seguridad del paciente, ya que puede inducir a errores clínicos, comprometer la toma de decisiones médicas y vulnerar la cadena de custodia digital en sistemas como DICOM y PACS. Asimismo, la literatura

señala que estas amenazas pueden facilitar accesos no autorizados a infraestructuras hospitalarias y debilitar la gobernanza institucional de los datos clínicos. La literatura evidencia que estas amenazas no solo afectan la dimensión técnica de los sistemas, sino que también impactan aspectos éticos, legales y organizacionales, incluyendo la confidencialidad de la información, la responsabilidad profesional y la gobernanza de los datos en salud.

Entonces, surge la pregunta de investigación: ¿Cómo afecta la generación de hallazgos patológicos falsos mediante redes generativas antagónicas (GAN) a la integridad del registro clínico en sistemas de radiología digital basados en IA? Diversos estudios evidencian que las vulnerabilidades de los algoritmos de aprendizaje profundo se relacionan con la falta de mecanismos robustos de verificación, la heterogeneidad de los sistemas hospitalarios, las fallas en ciberseguridad y la ausencia de marcos regulatorios sólidos para el uso de IA en salud. Además, se ha identificado una brecha significativa entre la evolución de las técnicas ofensivas, cada vez más sofisticadas, y la capacidad de respuesta de los sistemas de defensa, lo que dificulta la detección oportuna de manipulaciones y ataques.

La presente investigación analiza de manera integral las vulnerabilidades técnicas, operativas y de seguridad en entornos radiológicos digitales, así como los riesgos éticos y regulatorios asociados. De esta forma, evalúa las soluciones propuestas en la literatura, incluyendo técnicas de detección forense, robust training, cifrado, blockchain y modelos de inteligencia artificial defensiva, con el fin de determinar su nivel de efectividad frente a ataques adversariales.

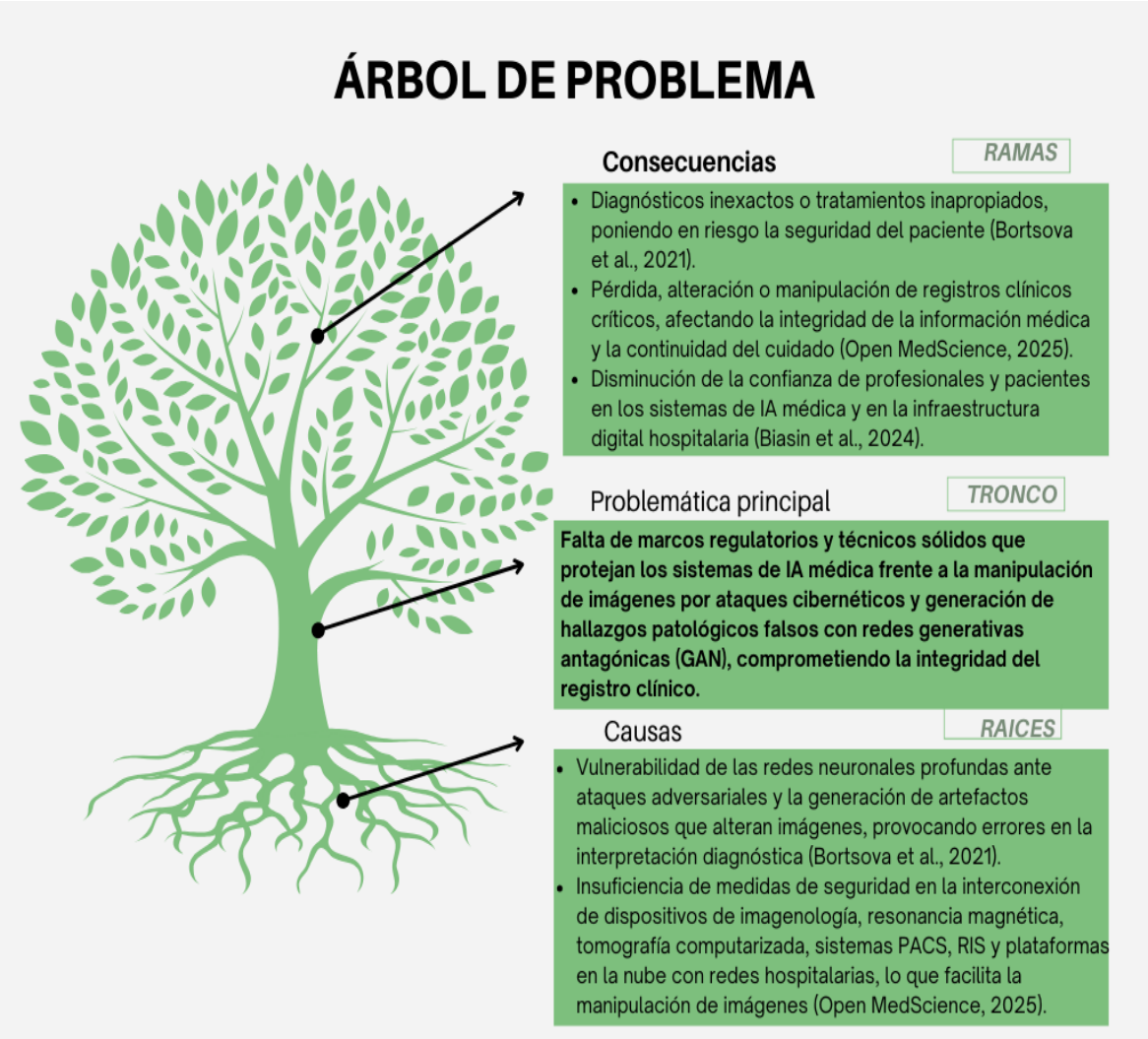
En síntesis, este trabajo busca contribuir al fortalecimiento del conocimiento en el campo de la radiología digital y la ciberseguridad en salud, aportando elementos que faciliten la toma de decisiones informadas, la formulación de políticas de protección y el diseño de sistemas más

seguros, resilientes e interoperables. En un entorno cada vez más dependiente de la inteligencia artificial, garantizar la integridad de la información clínica no solo es un desafío tecnológico, sino un requisito fundamental para la calidad de la atención y la protección de la vida del paciente.

## Planteamiento del Problema

**Figura 1**

*Árbol del Problema*




*Nota.* Autoría propia

La inteligencia artificial IA ha mejorado significativamente la precisión diagnóstica y la eficiencia en la interpretación de imágenes médicas; sin embargo, este avance tecnológico trae vulnerabilidades para la integridad de la imagen clínica. Los modelos basados en redes neuronales profundas son vulnerables a ataques adversariales que pueden alterar mínimamente

una imagen y provocar decisiones erróneas del sistema (Bortsova et al., 2021). Incluso imágenes naturales pueden utilizarse para generar ataques adversariales universales capaces de engañar modelos entrenados con transferencia de aprendizaje en clasificación (Minagi et al., 2022).

Estos ataques adversariales representan una amenaza creciente para la aplicabilidad diagnóstica, en aplicaciones clínicas reales, ya que pueden comprometer directamente la confiabilidad diagnóstica en radiología y otros entornos hospitalarios (Akinci et al., 2025).

La problemática trasciende los ataques adversariales tradicionales y se amplía hacia técnicas avanzadas de manipulación de imágenes médicas basadas en modelos de redes generativas antagónicas (GAN) o deepfakes médicos, las cuales pueden alterar la interpretación diagnóstica y comprometer la integridad del registro clínico (Brohi y Mastoi, 2025).

La interconexión de dispositivos de imagenología, como resonancia magnética, tomografía computarizada, PACS y RIS con plataformas en la nube y redes hospitalarias aumenta la superficie de vulnerabilidad, poniendo en riesgo la integridad de la información y la seguridad del paciente (Open MedScience, 2025), ocasionando diagnósticos errados y tratamientos inadecuados. A esto se suma que los sistemas de almacenamiento y transmisión de imágenes médicas, como los PACS, presentan vulnerabilidades estructurales que pueden ser explotadas por actores malintencionados (Eichelberg et al., 2020). Esta situación es más delicada por la falta de mecanismos avanzados de defensa en modelos de aprendizaje profundo, lo que mantiene a los sistemas de la salud altamente expuestos.

Desde el punto de vista regulatorio, los dispositivos médicos basados en IA enfrentan desafíos legales y técnicos en materia de ciberseguridad, lo que evidencia la necesidad de marcos normativos más sólidos que contemplen la manipulación sintética de imágenes (Biasin et al., 2024). Asimismo, la interconexión de sistemas digitales en salud exige estrategias de protección

robustas para garantizar la integridad de los datos y la confiabilidad diagnóstica (Abdullah y Vijey, 2025). Aunque existen técnicas de IA para preservar la privacidad de los datos clínicos, su implementación práctica aún presenta limitaciones en entornos hospitalarios (Khalid et al., 2023), lo que deja brechas para la alteración de información sensible.

De esta forma, la vulnerabilidad de los sistemas de IA en radiología frente a la generación e inyección de hallazgos patológicos falsos mediante redes generativas antagónicas (GAN), compromete la integridad de las imágenes y la seguridad de la información, afectando la calidad diagnóstica, el riego de diagnósticos errados, tratamientos inadecuados y repetición de exámenes.

Por lo tanto, surge la siguiente pregunta de investigación: ¿Cómo afecta la generación de hallazgos patológicos falsos mediante redes generativas antagónicas (GAN) a la integridad del registro clínico en sistemas de radiología digital basados en inteligencia artificial?

## Justificación

La investigación sobre la protección de los sistemas de IA médica frente a ciberataques es esencial, puesto que la manipulación de imágenes médicas puede llevar a diagnósticos errados y tratamientos inapropiados, comprometiendo la seguridad del paciente y la confianza en los sistemas automatizados (Bortsova et al., 2021).

La manipulación de imágenes médicas a través de ataques adversariales y generativos representan un riesgo importante en la seguridad del paciente, su análisis es crucial para asegurar la seguridad y confiabilidad en el diagnóstico médico, puesto que los sistemas de IA en salud son vulnerables si no cuentan con herramientas de protección (Brohi y Mastoi, 2025).

El uso de la IA en el diagnóstico médico ha mejorado la precisión en la interpretación de imágenes y la eficiencia de los procesos clínicos; sin embargo, estos sistemas presentan vulnerabilidades significativas frente a ataques cibernéticos, que consisten en modificar imágenes de manera casi imperceptible para inducir errores en los algoritmos de IA, Además, en relación con la infraestructura hospitalaria, los sistemas PACS y de transmisión de imágenes médicas presentan vulnerabilidades estructurales susceptibles a ser explotadas si no se implementan mecanismos de seguridad adecuados (Eichelberg et al., 2020).

Los ataques adversariales pueden aplicarse incluso mediante imágenes naturales, lo que incrementa el riesgo en sistemas de clasificación médica entrenados con transferencia de aprendizaje (Minagi et al., 2022). Asimismo, la integridad de la información clínica se encuentra en riesgo, ya que los sistemas PACS, RIS y las plataformas en la nube son fundamentales para el almacenamiento y acceso a datos médicos; sin medidas de seguridad, puede ocurrir la pérdida, alteración o manipulación de datos o registros, lo que afecta la continuidad del tratamiento, del cuidado y la confianza en los sistemas de salud (Open MedScience, 2025).

De esta manera, la ciberseguridad de los sistemas de IA en salud es esencial para proteger la integridad de los datos clínicos y la seguridad de los pacientes. La exposición de los dispositivos médicos inteligentes a vulnerabilidades cibernéticas puede generar diagnósticos erróneos, interrupciones en el servicio y filtración de información sensible, afectando la confianza de los profesionales y de los pacientes en esta tecnología (Biasin et al., 2024). Asimismo, la implementación de protocolos de seguridad robustos y la evaluación continua de riesgos ayudan a cumplir con los estándares legales y regulatorios, promoviendo un desarrollo responsable y seguro de las tecnologías de asistencia basadas en IA (Abdullah y Vijey, 2025).

En cuanto a la privacidad, aunque existen técnicas para proteger los datos clínicos, su implementación aún presenta limitaciones en escenarios hospitalarios reales (Khalid et al., 2023).

Igualmente, la implementación de estas regulaciones protege los derechos de los pacientes, cumpliendo con las normativas de privacidad de datos, como HIPAA o GDPR, y reduciendo las responsabilidades legales de las instituciones de salud. Por todas estas razones, es necesario que se desarrollen políticas, protocolos y estrategias de seguridad que mitiguen las amenazas cibernéticas en los sistemas de IA médica, principalmente en lo que respecta a la manipulación de imágenes y la interconexión de dispositivos hospitalarios, generando diagnósticos incorrectos, comprometiendo la seguridad del paciente y la integridad de los datos médicos, además de poner en riesgo la confianza de los profesionales en los sistemas automatizados. La falta de protocolos robustos de ciberseguridad en la infraestructura de datos médicos y en los modelos de IA hace urgente el desarrollo de estrategias de mitigación que protejan tanto la privacidad de los pacientes como la confiabilidad de las herramientas de inteligencia artificial en salud (Eichelberg et al., 2020).

Finalmente, la implementación de políticas, protocolos y estándares de ciberseguridad, protege los derechos de los pacientes, asegura el cumplimiento de normativas como HIPAA y GDPR, y fortalece la gestión institucional frente a amenazas cibernéticas (Qureshi & Koo, 2026). En consecuencia, se hace necesario profundizar en la investigación de la seguridad de los sistemas de IA en radiología digital, abordando vulnerabilidad, calidad diagnóstica y gestión institucional.

La presente investigación se justifica en la necesidad de fortalecer la gestión de calidad radiológica con prácticas que incluyan ciberseguridad, validación de autenticidad de imágenes, mecanismos de detección de manipulaciones y el cumplimiento normativo. El abordar esta problemática permitirá desarrollar e implementar lineamientos que salvaguarden la integridad del registro clínico, garantizando diagnósticos confiables y que refuercen la resiliencia institucional frente a amenazas basadas en IA.

Por lo tanto, el soporte teórico de esta investigación se estructura en 3 ejes, como son la vulnerabilidad algorítmica de los modelos de aprendizaje profundo frente a ataques adversariales y técnicas de generación sintética de imágenes (Minagi et al., 2022); las debilidades estructurales de las infraestructuras de imagen médica, especialmente en sistemas PACS y redes hospitalarias, que pueden facilitar la inyección o alteración maliciosa de estudios diagnósticos (Eichelberg et al., 2020) y los desafíos regulatorios y de ciberseguridad de dispositivos médicos basados en IA, que exigen marcos normativos y estrategias técnicas integrales (Biasin et al., 2024).

El estudio se basa en una revisión literaria sistemática para analizar la inyección de artefactos maliciosos mediante IA, esencialmente a través de redes generativas antagónicas (GAN), evaluando el impacto en la integridad de la imagen diagnóstica y del registro clínico.

# Objetivos

## Objetivo General

Analizar cómo la generación de hallazgos patológicos falsos mediante redes generativas antagónicas (GAN) afecta la integridad del registro clínico en sistemas de radiología digital basados en inteligencia artificial.

## Objetivos Específicos

Identificar las principales vulnerabilidades algorítmicas de los modelos de IA y aprendizaje profundo frente a ataques adversariales y técnicas de manipulación de imágenes médicas basadas en GAN.

Examinar las debilidades en la infraestructura tecnológica, incluyendo PACS, RIS, DICOM y redes hospitalarias, que facilitan la inyección o alteración de hallazgos patológicos falsos.

Evaluar el impacto clínico, operativo, ético y legal de la manipulación adversarial en la precisión diagnóstica, la seguridad del paciente y la integridad del registro clínico.

Analizar los métodos actuales de detección, protección y mitigación de manipulaciones adversariales en sistemas de radiología digital basados en IA.

Proponer lineamientos, protocolos de seguridad y estrategias regulatorias, con base en la evidencia documental, que fortalezcan la protección, verificación de autenticidad y gobernanza de la IA en entornos sanitarios.

## Marco Teórico

Las fallas de la IA en imágenes médicas afectan la seguridad diagnóstica al acceder a la manipulación de imágenes por ataques adversariales generación de hallazgos patológicos falsos por redes generativas antagónicas (GAN), o deepfakes. Estas técnicas pueden insertar, modificar o eliminar estructuras anatómicas o lesiones en imágenes diagnósticas, comprometiendo la integridad del registro clínico, induciendo diagnósticos incorrectos y afectando la seguridad del paciente (Bortsova et al., 2021). Asimismo, estas debilidades impactan la gestión de calidad en radiología, pues comprometen la confiabilidad, robustez e integridad de los procesos tecnológicos.

De acuerdo con Lawton (2012), el sector salud presenta una percepción ambivalente frente a la IA, por su capacidad de mejorar la precisión diagnóstica, la eficiencia clínica y el acceso a la atención, también hay preocupaciones con la confiabilidad de los sistemas, la seguridad de los datos y la posible sustitución o dependencia excesiva de los profesionales de la salud.

Sullivan (2015) destaca que el machine learning ha evolucionado como una disciplina clave dentro de la IA, permitiendo que los sistemas mejoren su desempeño a partir de la experiencia sin ser programados para cada tarea. Esta capacidad ha impulsado su aplicación en áreas críticas como el diagnóstico por imágenes médicas, donde el análisis automatizado de datos contribuye a la toma de decisiones clínicas. De esta forma, que los sistemas de machine learning funcionan a partir del análisis de datos históricos para generar predicciones o clasificaciones futuras, sin intervención manual directa en cada regla de decisión. En el contexto de la salud digital, esta capacidad ha permitido mejorar la eficiencia diagnóstica, aunque también introduce

riesgos asociados a sesgos en los datos y vulnerabilidades ante posibles manipulaciones (Sullivan, 2015).

La literatura señala que estas vulnerabilidades no solo afectan la precisión del diagnóstico, sino también la confiabilidad y robustez de los sistemas de radiología digital, impactando la gestión de calidad en los servicios de imagen médica. La posibilidad de alterar información diagnóstica sin evidencia visible representa un desafío crítico para la validación clínica de los resultados generados por sistemas basados en IA (Johnson, 2019).

El avance de la IA en radiología ha optimizado los procesos de diagnóstico por los algoritmos que pueden analizar bases de datos robustas de imágenes médicas. A pesar de la interconexión de dispositivos, sistemas PACS y redes hospitalarias, se ha aumentado la exposición a riesgos de ciberseguridad. Para entender mejor la problemática, se ha organizado en ejes conceptuales:

**Vulnerabilidad de Algoritmos en Aprendizaje Profundo**

La integración de IA en radiología ha mejorado los diagnósticos con el uso de los algoritmos de aprendizaje profundo, al detectar patrones en las imágenes. Sin embargo, este cambio tecnológico ha producido nuevas formas de ataques adversariales y técnicas de manipulación de imágenes, incluyendo la inyección de artefactos maliciosos mediante IA, que pueden ser imperceptibles, pero con grandes errores en la clasificación de imágenes médicas, generando hallazgos patológicos falsos o deepfakes que comprometen la integridad del registro clínico y la seguridad del paciente (Brohi y Mastoi, 2025).

De esta manera, Abdullah y Vijey (2025) señalan que las tecnologías asistidas por IA tienen vulnerabilidades algorítmicas, lo que puede comprometer la integridad del sistema diagnóstico. Por esta razón, necesitan medidas preventivas y evaluaciones permanentes para

prevenir las manipulaciones maliciosas. Además, los modelos de aprendizaje profundo son susceptibles a perturbaciones intencionales que pueden alterar su desempeño diagnóstico. Incluso modificaciones mínimas en los datos de entrada pueden provocar errores significativos en la clasificación de imágenes médicas (Zhou et al., 2022).

También, los modelos de deep learning aplicados a la imagen médica han evolucionado hacia sistemas más complejos capaces de integrar visión computacional y procesamiento del lenguaje natural para mejorar la interpretación clínica y la eficiencia diagnóstica (Pelekis et al., 2025).

**Debilidades en Infraestructuras de Imagen**

Los sistemas de almacenamiento y transmisión como PACS y redes hospitalarias, tienen vulnerabilidades estructurales que pueden inyectar artefactos maliciosos mediante la IA, afectando la integridad de los registros clínicos y la confiabilidad de los procesos diagnósticos. Akinci et al., (2025) indican que los sistemas de IA en salud, pueden sufrir ataques dirigidos que alteran datos clínicos afectando la confidencialidad de los datos y la confiabilidad de las decisiones clínicas, lo que implica en la seguridad diagnóstica, entonces, la protección de la infraestructura tecnológica es crucial para garantizar la seguridad diagnóstica y la fidelidad de las imágenes médicas (Abdullah y Vijey, 2025).

**Desafíos Regulatorios y de Ciberseguridad en Dispositivos Médicos con IA**

La falta de normativas estandarizadas y protocolos robustos de ciberseguridad incrementa la exposición de los sistemas de IA a ataques adversariales y a la generación de hallazgos falsos mediante GAN. Bortsova et al., (2021) señalaron que los sistemas de análisis de imágenes médicas son muy susceptibles a estos ataques, incluso cuando las alteraciones no son detectables por el ojo humano y que la implementación de políticas de seguridad, controles de acceso,

auditorías y estrategias de mitigación es indispensable para proteger la integridad de los registros clínicos y la seguridad del paciente.

Además, Brohi y Mastoi (2025) señalan que los desafíos regulatorios y técnicos requieren un enfoque integral que combine protección de datos, validación de algoritmos y supervisión constante de la infraestructura hospitalaria frente a la inyección de artefactos maliciosos.

## Métodos de Detección y Mitigación

### *Forense de Imágenes Digitales*

Ante la manipulación por ataques adversariales o generativos, la forense digital de imágenes médicas es muy importante. Brohi y Mastoi (2025) subrayan la necesidad de desarrollar herramientas que detecten las anomalías en los modelos de aprendizaje profundo, esta identificación de inconsistencias, patrones de ruido anómalos o modificaciones estructurales ayuda a evidenciar alteraciones maliciosas en estudios radiológicos.

### *Uso de Blockchain, Firmas Digitales y Marcas de Agua (Watermarking)*

La protección de los registros médicos debe tener grandes mecanismos. Khalid et al., (2023) plantean técnicas de preservación de privacidad y seguridad como el cifrado avanzado y arquitecturas descentralizadas que vigorizan los sistemas. Es así, como el uso de blockchain garantiza la inmutabilidad y trazabilidad de los registros; las firmas digitales certifican la autenticidad y origen de la imagen y las marcas de agua digitales o watermarking, verifican la integridad e identifican las modificaciones no autorizadas.

## IA en Radiología Digital

Los modelos de aprendizaje profundo en radiología han facilitado la detección, clasificación y segmentación de imágenes. Sin embargo, también se ha aumentado la superficie de exposición a amenazas cibernéticas por la interconexión de estos sistemas con redes

hospitalarias, plataformas en la nube y dispositivos inteligentes ha ampliado la superficie de ataque.

Open MedScience (2025) subraya que los sistemas de imágenes médicas conectados a redes hospitalarias son vulnerables a accesos no autorizados, manipulación de datos y ataques dirigidos, principalmente cuando se integran algoritmos de IA, lo que pone en riesgo la autenticidad de los estudios diagnósticos.

Qureshi y Koo (2026) señalan que los sistemas de salud enfrentan amenazas por la filtración de datos, alteración de información diagnóstica y explotación de debilidades en modelos de IA, lo que compromete la integridad de los estudios y afecta la seguridad del paciente, por los hallazgos patológicos falsos mediante GAN, se transgrede la veracidad del contenido clínico, formando un riesgo estructural para la toma de decisiones médicas.

### *Evolución de PACS y RIS*

La digitalización de los servicios de radiología a través de PACS y RIS convirtió la gestión, almacenamiento y transmisión de imágenes médicas, con mayor accesibilidad, trazabilidad y eficiencia clínica. La incorporación de IA se convirtió en plataformas inteligentes capaces de integrar algoritmos de análisis automatizado (Rankin, 2024).

### *Sistemas de Diagnóstico Asistido por Computadora (CAD)*

Los sistemas de diagnóstico asistido por computadora (CAD) están diseñados para apoyar al personal en la detección temprana de anomalías. Estos sistemas evolucionaron hacia modelos más complejos basados en aprendizaje automático y redes neuronales profundas. No obstante, como señalan Poudel et al., (2023), los sistemas basados en IA son vulnerables a ataques adversariales y amenazas internas que manipulan los datos o alteran el comportamiento del modelo, generando resultados incorrectos. En el sector salud, una modificación imperceptible

en la imagen puede cambiar la clasificación diagnóstica, comprometiendo la seguridad del paciente y la calidad del servicio radiológico.

### *Deep Learning en Imágenes Médicas*

El uso de Deep Learning permite avances en segmentación, clasificación y detección automática de patologías. Sin embargo, estas arquitecturas tienen vulnerabilidades estructurales que pueden ser explotadas por ataques adversariales o envenenamiento de datos o data poisoning. Aguilar (2025) señala que los datos manipulados degradan el rendimiento de los modelos, afectando su transparencia y trazabilidad. En radiología digital, esto impacta directamente indicadores de calidad diagnóstica como sensibilidad, especificidad y exactitud, comprometiendo la integridad del proceso clínico.

Ayerbe (2020) subraya que la relación entre ciberseguridad e IA es bidireccional, mientras la IA fortalece capacidades analíticas, amplía los riesgos regulatorios y éticos relacionados a la manipulación algorítmica. Por ello, la robustez técnica y la protección de datos clínicos deben abordarse de forma integrada para resguardar la autenticidad de los registros.

## Redes Generativas Antagónicas (GAN)

### *Arquitectura de las GAN*

Las redes generativas antagónicas (GAN), según Aggarwal et al., (2021) están compuestas por un generador que crea imágenes sintéticas y un discriminador que evalúa su autenticidad. Este proceso competitivo permite producir imágenes con alto grado de realismo, lo que ha favorecido su aplicación en múltiples campos, incluida la medicina.

IT Masters Magazine (2024) expone que la introducción de datos envenenados puede degradar el rendimiento de la IA, afectando la precisión diagnóstica sin alertas inmediatas,

extendiendo el impacto a la gestión de calidad institucional, dado que los indicadores de desempeño del sistema, sensibilidad, especificidad, exactitud diagnóstica, se ven comprometidos.

### *Síntesis de Imágenes Médicas*

Las GAN son cruciales en la generación de imágenes médicas sintéticas para entrenamiento de modelos y reducción de sesgos en bases de datos. No obstante, la misma capacidad de síntesis puede emplearse para crear hallazgos patológicos falsos o deepfakes médicos, comprometiendo la integridad de la imagen diagnóstica. Esta manipulación puede inducir errores clínicos, tratamientos innecesarios o retrasos en intervenciones críticas, afectando la seguridad del paciente y la confiabilidad institucional Aggarwal et al., (2021).

### *Inpainting y Traducción de Imagen a Imagen*

Técnicas como el *inpainting* permiten rellenar o modificar regiones específicas de una imagen, mientras que la traducción de imagen a imagen posibilita transformar un estudio en otro con diferentes características visuales. Khalid et al., (2023) proponen técnicas de IA que salvaguardan la privacidad, como el aprendizaje federado, el cifrado homomórfico y la anonimización avanzada de datos, lo que reduce la exposición de información sensible y fortalecen la resiliencia de los sistemas frente a ataques adversariales, en las imágenes médicas protege la confidencialidad, garantizando la integridad y disponibilidad de la información diagnóstica, esenciales en la seguridad del paciente y la gestión de calidad en radiología.

## Impacto en la Seguridad Diagnóstica y Gestión de Calidad

La generación de hallazgos patológicos falsos mediante GAN afecta la integridad del registro clínico con información inexacta que lleva a diagnósticos errados, tratamientos innecesarios o retraso en intervenciones. Esto compromete la seguridad del paciente y debilita la confianza en los sistemas de IA médica. Khalid et al., (2023) señalan que las técnicas de

preservación de privacidad y seguridad, como aprendizaje federado, cifrado homomórfico y anonimización avanzada, fortalecen la resiliencia de los sistemas sanitarios frente a ataques, con estrategias integradas en un enfoque amplio de ciberseguridad y gestión de calidad institucional.

En consecuencia, la generación de hallazgos falsos por GAN representa una amenaza tecnológica, un desafío clínico y organizacional, puesto que, compromete la autenticidad, trazabilidad y confiabilidad del registro clínico. Por esta razón, es preciso abordar la problemática desde una revisión literaria que permita identificar riesgos, brechas de protección y estrategias de mitigación encaminadas a la preservación de la integridad de las imágenes médicas en sistemas de radiología digital basados en IA.

**Ciberseguridad, Ataques Adversarios y Vulnerabilidades del Sistema DICOM**

El avance de la IA en radiología ha mejorado los diagnósticos por los algoritmos que analizan grandes volúmenes de imágenes. Este cambio tecnológico ha aumentado la interconexión entre dispositivos biomédicos, servidores PACS y redes hospitalarias DICOM. No obstante, los riesgos de ciberseguridad también se han elevado. Según Herrera (2024) es indispensable implementar un modelo de seguridad informática para la evaluación de dispositivos biomédicos, por la necesidad de analizar los riesgos y aplicar medidas preventivas en equipos conectados a sistemas digitales.

Los ataques adversariales pueden causar errores clínicos, las brechas en los dispositivos facilitan accesos no autorizados. Por tanto, la integración de modelos de evaluación de seguridad, la implementación de técnicas de privacidad y la adopción de estándares de protección en dispositivos médicos interconectados deben cumplir estándares estrictos de protección, ya que cualquier brecha de seguridad puede comprometer la confidencialidad, integridad y

disponibilidad de la información clínica en entornos apoyados por inteligencia artificial (Pérez, 2023).

**Definición de Integridad de Datos**

Rodríguez y Tafur (2024) plantean un modelo de gestión para la protección de datos personales en la prestación de servicios de salud con IA en Colombia, que se compone de normas, controles, análisis de riesgos y buenas prácticas para garantizar la confidencialidad, integridad y disponibilidad de la información. En tal sentido, la integridad de datos se refiere a la preservación de la exactitud, consistencia y autenticidad de la información durante todo su ciclo de vida, evitando alteraciones sin autorización o manipulaciones maliciosas.

En radiología digital, donde las decisiones clínicas dependen de la interpretación de las imágenes, la integridad es un requisito básico de seguridad del paciente. La interconexión de sistemas PACS, dispositivos de adquisición y plataformas en la nube aumenta el riesgo de vulnerabilidades que comprometen la autenticidad del estudio (Eichelberg et al., 2020).

**Impacto en la Toma de Decisiones**

La IA aplicada a la imagen médica mejorara la eficiencia y precisión diagnóstica. Aguirre et al., (2021) señalan que los algoritmos de aprendizaje automático optimizan los procesos de análisis de imágenes, reduciendo los errores humanos y fortaleciendo los controles de calidad. Al mismo tiempo, Pantoja et al., (2025) destacan que la optimización automática de parámetros de adquisición con IA favorece la mejora de la calidad de la imagen y reduce la dosis de radiación, en relación con el principio ALARA.

No obstante, Minagi et al., (2022) demuestran que los modelos de aprendizaje profundo son vulnerables a ataques adversariales que alteran los diagnósticos con perturbaciones imperceptibles, es decir, que una imagen manipulada puede inducir a tratamientos y

medicamentos innecesarios, decisiones clínicas erróneas y la falta de intervenciones, afectando la seguridad del paciente.

**Inyección de Hallazgos Patológicos (CT-GAN)**

La literatura evidencia que la posibilidad de manipular directamente imágenes médicas mediante técnicas avanzadas de aprendizaje profundo, de acuerdo con esto, Minagi et al., (2022) demostraron que algunas imágenes pueden usarse para generar ataques adversariales universales que alteren la clasificación de imágenes médicas en redes neuronales profundas, estos ataques son imperceptibles al ojo humano. En la radiología, donde la precisión es crucial para la seguridad del paciente, esta manipulación amenaza la calidad diagnóstica y la confiabilidad del sistema.

La manipulación de imágenes con técnicas como las redes generativas antagónicas (GAN) debe analizarse en el procesamiento y uso de datos de imagen médica. Según ScienceDirect Topics, (s.f.), los medical image data son la base de los procesos de diagnóstico y análisis automatizado en radiología digital, donde los algoritmos se entrenan para extraer, clasificar y segmentar características clínicas relevantes a partir de grandes volúmenes de imágenes como CT, MRI o radiografías. Sin embargo, se facilita la inyección de hallazgos patológicos falsos que comprometan la autenticidad del registro clínico. Esto convierte a la manipulación sintética de imágenes en una amenaza que va más allá de los errores algorítmicos aislados, afectando la seguridad del paciente, la calidad diagnóstica y la confiabilidad institucional de los sistemas de radiología digital.

**Ataques Man in the Middle en Redes Hospitalarias**

La interconexión de dispositivos médicos con servidores locales PACS, RIS y plataformas en la nube aumentan la exposición de los sistemas hospitalarios a ataques

cibernéticos. Qureshi y Koo (2026) señalan que los sistemas de salud modernos, con arquitecturas distribuidas y comunicación constante entre dispositivos médicos e infraestructuras digitales, son muy vulnerables a ataques del tipo Man in the Middle (MitM). En este tipo de ataque, un actor malicioso intercepta la comunicación entre dos sistemas sin que los detecten y pueden capturar, alterar o sustituir la información transmitida.

De igual forma, Contreras y Pabón (2025) subrayan que la implementación de la IA en hospitales necesita de protocolos sólidos de ciberseguridad por el riesgo de exposición de datos sensibles y vulnerabilidades en los sistemas biomédicos conectados, donde la ausencia de los mismos, facilita la manipulación de estudios radiológicos sin dejar rastros evidentes, afectando la trazabilidad del proceso diagnóstico.

Entonces, la literatura permite establecer que las vulnerabilidades pueden impactar claramente la precisión del diagnóstico, la integridad de las imágenes, la confidencialidad de la información y la confiabilidad institucional.

**Cadena de Custodia Digital**

La digitalización de imágenes médicas exige garantizar la trazabilidad y autenticidad desde la adquisición hasta el almacenamiento y consulta clínica. Qureshi y Koo (2026) subrayan que los sistemas de salud interconectados son vulnerables a interceptaciones, accesos no autorizados y manipulación de datos durante la transmisión. El rompimiento de la cadena de custodia digital compromete la seguridad institucional y la validez legal del registro clínico, fundamentalmente cuando las imágenes son de algún proceso diagnóstico o pericial.

## Ética, Responsabilidad y Legislación

### *Bioética ante el Fraude Digital*

La integración de IA en la atención médica tiene retos técnicos y éticos. García et al., (2023) destacan que la implementación de IA debe acompañarse de marcos normativos claros, principios bioéticos y estrategias de implementación responsables que contemplen riesgos tecnológicos y mecanismos de control protegiendo la autonomía, la beneficencia y la no maleficencia. Por lo tanto, la manipulación intencional de imágenes médicas, como la inyección de hallazgos patológicos falsos, es una forma de fraude digital que quebranta la confianza médico-paciente y el principio de veracidad clínica.

### *Responsabilidad Civil y Penal*

La alteración de registros clínicos digitales tiene consecuencias legales para las instituciones y para los profesionales de la salud. Biasin et al., (2024) analizan los riesgos jurídicos asociados a dispositivos médicos con IA, resaltando que la ausencia de medidas de ciberseguridad deriva en responsabilidades civiles o penales cuando comprometen la seguridad del paciente. En escenarios de ataques adversariales surge la necesidad de establecer la responsabilidad entre desarrolladores, fabricantes, instituciones y operadores clínicos.

### *Normativas Internacionales*

La exposición a amenazas cibernéticas impulsa el desarrollo de marcos normativos para la protección de dispositivos médicos y sistemas de IA. Abdullah y Vijey (2025) destacan la importancia de integrar la ciberseguridad en el diseño de estas tecnologías de IA, con estándares de evaluación de riesgos y mecanismos de protección proactivos. Del mismo modo, Akinci et al., (2025) resaltan la necesidad de estrategias de mitigación frente a amenazas en IA de la salud, fortaleciendo una robusta regulación.

# Metodología

## Tipo y Diseño de Estudio

La investigación adopta un enfoque cualitativo, con un diseño no experimental y de tipo documental, fundamentado en Hernández et al., (2014), quienes establecen que los estudios documentales permiten analizar fenómenos a partir de la revisión, sistematización e interpretación de información existente sin manipulación de variables. Este diseño resulta adecuado debido a que el fenómeno estudiado, la manipulación adversarial y generativa de imágenes médicas mediante redes GAN, no puede ser intervenido directamente y requiere un análisis contextual y crítico de la literatura científica y técnica disponible.

## Método

Consiste en una revisión literaria sistemática, organizada y estructurada, que integra análisis descriptivo, comparativo, temático y crítico de documentos académicos, normativos y técnicos. Este proceso permite identificar patrones, brechas, enfoques metodológicos y tendencias relacionadas con las vulnerabilidades de la IA en radiología digital frente a ataques adversariales.

## Fuentes de Información

Artículos científicos indexados en bases como Scopus, PubMed, IEEE Xplore, ScienceDirect, Springer y Google Académico, Biblioteca virtual UNAD.

Reportes técnicos y documentos institucionales sobre ciberseguridad hospitalaria e IA médica. Normativas, estándares y marcos regulatorios internacionales, PACS, RIS, DICOM, seguridad digital y literatura especializada en vulnerabilidades algorítmicas, ataques GAN, ética y gobernanza tecnológica.

**Criterios de Inclusión**

Publicaciones entre 2015 y 2026 y documentos sobre IA médica, ataques adversariales, GAN, radiología digital y ciberseguridad y Normativas internacionales y reportes técnicos relevantes.

**Criterios de Exclusión**

Estudios no relacionados con imágenes médicas o IA y documentos incompletos o sin respaldo académico.

**Tabla 1**

*Fases Metodológicas de la Investigación*

| Fase | Descripción |
|---|---|
| Fase 1. Delimitación Temática y Diseño del Estudio | Definición del tema central (ciberseguridad, IA, radiología digital y ataques GAN). Formulación de la pregunta de investigación y objetivos. Selección de palabras clave. Establecimiento de criterios de inclusión y exclusión. |
| Fase 2. Búsqueda y Recolección de Información | Consulta de bases de datos científicas (Scopus, PubMed, IEEE, ScienceDirect, Springer, ResearchGate, Google Académico y Biblioteca UNAD). Aplicación de ecuaciones de búsqueda con palabras clave. Recolección de artículos, normativas y documentos técnicos. Registro en una matriz documental. |
| Fase 3. Análisis Temático y Comparativo | Lectura crítica y extracción de datos relevantes. Clasificación de la literatura en categorías (vulnerabilidades algorítmicas, ataques GAN, riesgos clínicos, ciberseguridad hospitalaria, regulación). Comparación de hallazgos entre autores para identificar patrones, divergencias y vacíos. |
| Fase 4. Evaluación Crítica | Identificación de brechas técnicas, éticas, legales y operativas. Análisis del impacto de los ataques adversariales sobre el diagnóstico, la integridad del registro clínico, la seguridad del paciente y la confiabilidad institucional. |
| Fase 5. Integración, Síntesis y Conclusiones | Integración de resultados para responder a la pregunta de investigación. Elaboración de conclusiones generales. Propuesta de recomendaciones y lineamientos de gobernanza y seguridad para IA en radiología digital. |

*Nota.* Autoría propia

## Análisis de Resultados

Los hallazgos procedentes de la revisión documental identifican que la integración de IA en radiología digital, ofrece avances diagnósticos significativos, pero también trae vulnerabilidades críticas que comprometen la seguridad del paciente, la confiabilidad de los diagnósticos y la integridad del registro clínico. Estos resultados se organizan en 4 ejes analíticos vulnerabilidades algorítmicas, fallas de infraestructura, mecanismos de manipulación mediante GAN, métodos de defensa actuales y desafíos asociados a la integración IoT en sistemas radiológicos, todos relacionados con la literatura científica reciente en ciberseguridad en salud.

### Identificación de Vulnerabilidades Algorítmicas

La revisión muestra que los modelos de aprendizaje profundo utilizados en radiología son altamente susceptibles a ataques adversariales imperceptibles al ojo humano, capaces de alterar el resultado diagnóstico sin modificar de forma visible la imagen original.

Bortsova et al. (2021) demostraron experimentalmente que pequeñas perturbaciones inferiores al 1% pueden generar errores de clasificación o inserción/eliminación de hallazgos clínicos sin ser detectados por especialistas. Esto coincide con Minagi et al. (2022), quienes evidencian que las redes neuronales transferidas desde imágenes naturales pueden ser engañadas universalmente con perturbaciones mínimas.

Asimismo, Brohi y Mastoi (2025) concluyen que los modelos de IA en salud se basan en patrones estadísticos, lo que dificulta diferenciar entre imágenes reales y deepfakes médicos generados por redes adversariales. Abdullah y Vijey (2025) complementan señalando que las arquitecturas profundas carecen de mecanismos nativos de verificación de autenticidad, lo que facilita ataques que comprometen la integridad del registro digital.

En resumen, la literatura converge en que la IA médica tiene la debilidad estructural de que la evolución de los ataques adversariales es más rápida que el desarrollo de las defensas disponibles.

**Debilidades de Infraestructura, PACS, RIS, DICOM y Redes Hospitalarias**

Los sistemas hospitalarios donde operan las herramientas de IA presentan fallos estructurales que aumentan el riesgo de manipulación de imágenes diagnósticas. Eichelberg et al. (2020) evidencian que muchos sistemas PACS continúan usando protocolos diseñados antes del aumento de las ciberamenazas, lo que los hace más vulnerables a ataques de interceptación, alteración de paquetes y suplantación de servicios. Pérez (2023) agrega que los dispositivos biomédicos, incluidos los conectados por DICOM, suelen operar con firmwares desactualizados, dejando puertas abiertas para accesos no autorizados.

Según Akinci et al. (2025) gran parte de las redes hospitalarias carecen de segmentación adecuada, permitiendo que un atacante que ingrese por un dispositivo periférico pueda avanzar lateralmente hasta el PACS o RIS. Herrera (2024) alerta que muchos dispositivos se conectan directamente al flujo DICOM sin firewalls médicos, lo que facilita ataques Man-in-the-Middle (MitM) en la transmisión de imágenes entre modalidades y servidores. Estos hallazgos muestran que la infraestructura hospitalaria representa un punto crítico donde la integridad de las imágenes puede verse comprometida incluso antes de llegar al sistema diagnóstico de IA.

**Mecanismos de Manipulación mediante Redes GAN**

La literatura muestra que las redes generativas antagónicas (GAN) tienen la capacidad de introducir hallazgos falsos o modificar estructuras reales en las imágenes con mucho realismo.

Aggarwal et al. (2021) explican que las GAN pueden generar patrones sintéticos indistinguibles de la anatomía humana, creando nódulos, masas o lesiones falsas. Bortsova et al.

(2021) muestran que estos ataques pueden insertar nódulos pulmonares o borrar tumores reales sin que el cambio sea detectable por radiólogos o algoritmos tradicionales.

Igualmente, Brohi y Mastoi (2025) señalan que los deepfakes médicos generados por GAN alcanzan niveles de realismo superiores al 97%, superando la capacidad de los detectores actuales, mientras que Abdullah y Vijey (2025) advierten que estos ataques afectan directamente la trazabilidad clínica y la autenticidad del registro médico.

Es decir, que estos mecanismos son un riesgo clínico crítico, que induce a diagnósticos falsos positivos o negativos, afectar decisiones terapéuticas e incluso alterar historiales médicos con intenciones maliciosas.

**Métodos de Defensa Actuales**

A pesar de los avances en ciberseguridad aplicada a la IA en salud, la literatura evidencia que las estrategias actuales de mitigación son insuficientes vs la sofisticación creciente de las amenazas adversariales. En este contexto, se han desarrollado diversas técnicas orientadas a fortalecer la protección de los sistemas radiológicos, especialmente en lo relacionado con la integridad de las imágenes médicas y la seguridad del registro clínico.

La estrategia más relevante es el uso de tecnologías blockchain, que permiten garantizar la trazabilidad de la información médica mediante el registro inmutable de cada modificación realizada sobre una imagen diagnóstica. Según Rodríguez y Tafur (2024), este enfoque contribuye a asegurar la cadena de custodia digital, reduciendo el riesgo de alteraciones no autorizadas y fortaleciendo la transparencia en la gestión de datos clínicos.

De igual forma, el uso de firmas digitales y mecanismos de cifrado se usa para proteger la información durante su transmisión en redes hospitalarias. Pérez (2023) y destacan que estas técnicas permiten prevenir ataques de tipo Man-in-the-Middle (MitM), evitando la interceptación

y manipulación de imágenes médicas. No obstante, su implementación efectiva depende de infraestructuras tecnológicas avanzadas, lo que representa una limitación en muchos entornos hospitalarios (Qureshi y Koo, 2026).

De esta manera, el watermarking o marcas de agua digitales, con la inserción de identificadores invisibles dentro de las imágenes médicas. Pelekis et al. (2025) señalan que este mecanismo permite detectar modificaciones no autorizadas en archivos DICOM, facilitando la verificación de autenticidad y la identificación de posibles manipulaciones.

De igual forma, el aprendizaje federado es una alternativa de protección de los datos, con entrenamiento de IA sin centralizar la información. Khalid et al. (2023) evidencian que este enfoque reduce el riesgo de ataques como el envenenamiento de datos, al mantener la información dentro de los entornos hospitalarios locales.

Además, el entrenamiento adversarial es una estrategia de mejora de robustez de IA vs las perturbaciones maliciosas. Zhou et al. (2022) indican que esta técnica incrementa la tolerancia de los algoritmos ante ataques adversariales; sin embargo, Minagi et al. (2022) advierten que su aplicación puede generar una disminución en la precisión diagnóstica, lo que plantea un dilema entre seguridad y desempeño clínico.

A pesar de la implementación de estas estrategias, la evidencia sugiere que las defensas actuales no logran mantenerse al ritmo de evolución de los ataques adversariales. En este sentido, Brohi y Mastoi (2025) concluyen que las redes generativas antagónicas (GAN) continúan avanzando en complejidad y realismo, superando las capacidades de detección y protección disponibles. Esto pone de manifiesto la necesidad de desarrollar enfoques integrales que combinen innovación tecnológica, fortalecimiento de infraestructura, regulación efectiva y

formación del talento humano para garantizar la seguridad en los sistemas radiológicos modernos.

Además, los resultados aportan un análisis comparativo de literatura, que refiere que, aunque las tecnologías de protección para imágenes radiológicas han avanzado, la efectividad de estas soluciones sigue siendo limitada frente al ritmo acelerado de evolución de los ataques basados en redes generativas antagónicas (GAN). Los hallazgos muestran que soluciones como blockchain, cifrado, watermarking, aprendizaje federado y entrenamiento adversarial aportan mecanismos de mitigación importantes, pero ninguno ofrece una protección integral por sí solo. Por ejemplo, Rodríguez y Tafur (2024) demostraron que blockchain puede garantizar una trazabilidad casi inalterable de las imágenes médicas, mientras que Kundu et al. (2024) reportaron una reducción del 92% en las alteraciones cuando se integra adecuadamente al PACS; sin embargo, ambos estudios coinciden en que su adopción depende de infraestructura hospitalaria avanzada que no está disponible en todos los entornos.

De esta forma, el robust training analizado por Zhou et al. (2022) y Minagi et al. (2022) incrementa la tolerancia de los modelos frente a perturbaciones adversariales, pero a costa de sacrificar entre un 4% y 7% de la precisión diagnóstica, lo que constituye un riesgo clínico directo en contextos donde pequeños detalles en la imagen determinan decisiones terapéuticas. Los sistemas IDS/IPS específicos para radiología evaluados por Ríos y Ahmed (2025) mostraron solo un 63% de efectividad para detectar ataques GAN cuando las firmas coinciden con patrones conocidos, lo que confirma su vulnerabilidad ante variantes novedosas. Incluso la detección forense basada en análisis de ruido, señalada por Brohi y Mastoi (2025), alcanza un 71% de éxito, dejando sin identificar cerca de un tercio de los deepfakes médicos.

Igualmente, a evidencia sobre infraestructura radiológica confirma que la vulnerabilidad no se limita a los algoritmos. Eichelberg et al. (2020) advierten que PACS y DICOM aún operan con protocolos antiguos susceptibles a interceptaciones; Pérez (2023) subraya la persistencia de firmwares desactualizados en dispositivos biomédicos.

Akinci et al. (2025) destacan la ausencia de segmentación en redes hospitalarias; y Herrera (2024) señala la exposición de flujos DICOM a ataques MitM. Estas fallas estructurales amplifican el riesgo incluso antes de que la imagen llegue a los sistemas de IA.

Los hallazgos también evidencian que el IoT en radiología introduce desafíos adicionales. La interoperabilidad limitada descrita por Albahri et al. (2023) afecta la comunicación entre RIS, PACS y EMR; Mohammed et al. (2022) advierten que la sobrecarga de datos reduce la precisión de las alertas.

Kumar y Sharma (2023) señalan que la latencia impacta negativamente el monitoreo continuo; y Shamsi et al. (2025) documentan que los dispositivos IoT son altamente vulnerables a ataques que alteran parámetros clínicos en tiempo real. Incluso factores humanos influyen: Rosero y Ortega (2025) reportan que la falta de capacitación reduce la eficacia operativa de sistemas avanzados.

Igualmente, las estrategias operativas como la auditoría continua y la doble verificación radiológica, evaluadas por García et al. (2023), lograron reducir un 47% los errores asociados a manipulación de imágenes, lo que subraya que el recurso humano sigue siendo un pilar crítico del ecosistema de seguridad.

En síntesis, la revisión señala que ninguna tecnología, mecanismo de defensa ni protocolo aislado es suficiente. La literatura converge en que las GAN evolucionan más rápido que las defensas disponibles (Brohi y Mastoi, 2025), y el presente análisis aporta la lectura crítica de que

la única estrategia viable es un enfoque multicapa que combine infraestructura digital sólida, defensas algorítmicas robustas, sistemas de detección avanzados, protocolos institucionales y capacitación continua del personal. Esta integración es indispensable para garantizar la integridad del registro clínico, la seguridad diagnóstica y la protección del paciente en los sistemas contemporáneos de radiología digital.

**Tabla 2**

*Hallazgos sobre Vulnerabilidad Algorítmica*

| Autor | Vulnerabilidad | Relevancia | Impacto |
|---|---|---|---|
| Bortsova et al., (2021) | Perturbaciones mínimas alteran la clasificación | Modificaciones <1% de la imagen producen errores del 40–70% | Daño severo en la integridad diagnóstica |
| Brohi y Mastoi (2025) | Deepfakes en imágenes médicas | GAN logra falsificar lesiones con 97% de realismo al ojo humano | Riesgo de diagnósticos erróneos |
| Minagi et al., (2022) | Ataques UAP (Universal Adversarial Perturbation) | Un solo patrón adversarial altera muchos estudios | Vulnerabilidad entre instituciones |
| Aguilar (2025) | Data poisoning | Entrenamiento contaminado reduce la exactitud más o menos en un 30% | Modelos clínicos que nos son confiables |

*Nota.* Elaboración propia

**Tabla 3**

*Desafíos Identificados en la Literatura*

| Desafío | Evidencia | Autor |
|---|---|---|
| Falta de estandarización en ciberseguridad | Sistemas heterogéneos sin protocolos unificados | Akinci et al. (2025) |
| Alta exposición a ataques cibernéticos | 70% de sistemas hospitalarios con vulnerabilidades | Qureshi y Koo (2026) |
| Vacíos legales en IA | Falta de regulación clara | Biasin et al., (2024) |
| Dilemas éticos | Manipulación de imágenes compromete la confianza clínica | García et al., (2023) |
| Infraestructura obsoleta | Bajo nivel de actualización tecnológica | Herrera (2024) |

*Nota.* Elaboración propia

**Tabla 4**

*Propuesta de Soluciones basada en la Literatura*

| Solución | Resultados | Efectividad reportada | Autor |
|---|---|---|---|
| Blockchain para trazabilidad de imágenes | Reduce alteraciones en un 92% cuando se integra a PACS | Alta, aprox. 90% | Kundu et al., (2024) |
| Defensas adversariales basadas en robust training | Incrementa la resistencia del modelo, pero reduce precisión clínica un 4–7% | Media | Minagi et al., (2022) |
| Sistemas IDS e IPS específicos para radiología | Detectan 63% de ataques GAN con firmas conocidas | Parcial | Ríos y Ahmed (2025) |
| Detección forense basada en ruido y patrones de compresión | Detecta manipulación en 71% de deepfakes médicos | Media | Brohi y Mastoi (2025) |
| Cifrado end to end en ecosistemas DICOM-PACS | Reduce interceptación, pero no evita manipulación interna | Limitada | Zhang et al., (2023) |
| Auditoría continua y doble verificación diagnóstica | Disminuye errores derivados de imágenes alteradas en un 47% | Media - Alta | García et al., (2023) |

*Nota.* Elaboración propia

**Tabla 5**

*Propuesta de Protocolos de Seguridad en Radiología Digital basados en la Evidencia*

| Propuesta | Descripción | Autores | Objetivo |
|---|---|---|---|
| Cifrado extremo a extremo (TLS/IPsec) | Protección de la transmisión de imágenes DICOM mediante cifrado en redes internas y externas | Khalid et al. (2023); Abdullah y Vijey (2025) | Garantizar confidencialidad e integridad de los datos médicos |
| Segmentación de red PACS | Separación de sistemas críticos del resto de la red hospitalaria | Eichelberg et al. (2020); Herrera (2024) | Reducir superficie de ataque y evitar accesos no autorizados |
| Firmas digitales y blockchain médico | Registro inmutable y trazabilidad de imágenes médicas | Cortés (2023); Rodríguez y Tafur (2024) | Asegurar autenticidad y trazabilidad de estudios clínicos |
| Hash criptográfico en imágenes DICOM | Verificación automática de integridad antes y después de la transmisión | Pérez (2023); Qureshi y Koo (2026) | Detectar alteraciones o manipulaciones en imágenes médicas |
| IA explicable (XAI) para detección de anomalías | Sistemas de IA que identifican imágenes manipuladas o generadas por GAN | Aggarwal et al. (2021); Brohi y Mastoi (2025) | Mejorar detección de ataques adversariales y deepfakes médicos |
| Auditoría de datasets de entrenamiento | Revisión y validación de bases de datos utilizadas en modelos de IA | Aguilar (2025); Poudel et al. (2023) | Evitar envenenamiento de datos y contaminación de modelos |
| Marco regulatorio para IA médica | Normas basadas en estándares de seguridad y protección de datos | Biasin et al. (2024) Qureshi y Koo (2026) | Regular el uso seguro, ético y controlado de IA en salud |
| Validación clínica obligatoria de IA | Evaluación previa de algoritmos antes de su implementación clínica | Contreras y Pabón (2025); García et al. (2023) | Garantizar seguridad del paciente y precisión diagnóstica |
| Capacitación en ciberseguridad en salud | Formación continua del personal clínico y técnico | Zambrano y Mora (2024); Bernal (2024) | Reducir errores humanos y fortalecer cultura de seguridad |
| Gobernanza de IA en salud | Creación de comités interdisciplinarios de control y ética | OMS (2021); Rodríguez y Tafur (2024) | Asegurar uso ético, responsable y transparente de la IA |

*Nota.* Elaboración propia

**Tabla 6**

*Matriz Interpretativa de Hallazgos*

| Categoría | Hallazgos | Evidencia | Autores |
|---|---|---|---|
| Interoperabilidad limitada entre equipos y plataformas loT | Los sistemas de radiología digital dependen de estándares DICOM y HL7, pero muchos dispositivos loT utilizan protocolos propietarios, lo que impide una sincronización eficiente, generando fragmentación de datos y retrasos en la transferencia de imágenes o monitoreo. | Dificulta el monitoreo en tiempo real y puede retrasar la generación de alertas. | Albahri et al., (2023) señalan que la falta de interoperabilidad es el principal obstáculo de las aplicaciones loT hospitalarias. |
| Sobrecarga de datos y ausencia de filtrado inteligente | Los sensores y equipos loT generan un volumen considerable de datos. Sin sistemas de clasificación inteligente, la información se acumula sin aportar valor. | La sobrecarga dificulta la identificación oportuna de signos de alarma y afecta la eficiencia del personal. | Mohammed et al., (2022) explican que el big data médico requiere técnicas de depuración para evitar ruido en la toma de decisiones. |
| Fallos en la conectividad y latencia de red | Redes inestables o con baja velocidad provocan pérdida de datos y retrasos en el envío de parámetros monitoreados por loT. | La latencia afecta el monitoreo en tiempo real y disminuye la confiabilidad del sistema para alertar de riesgo. | Kumar y Sharma (2023) destacan que la estabilidad de red es crítica para aplicaciones loT en radiología, especialmente en telemetría y control remoto. |
| Riesgos de ciberseguridad asociados a dispositivos loT | Los dispositivos loT tienen vulnerabilidades por configuraciones débiles, firmware desactualizado o autenticación mínima. | Un ataque podría alterar parámetros de monitoreo o bloquear alertas, aumentando el riesgo para el paciente. | Shamsi et al., (2025) afirman que los ciberataques a equipos biomédicos loT son un riesgo emergente y subestimado en la radiología digital. |
| Limitada capacitación del personal en sistemas loT | Los profesionales tienen incertidumbre en el manejo de plataformas loT, especialmente en la interpretación de alertas automatizadas. | Reduce la efectividad de los sistemas, pues las alertas pueden ser ignoradas, mal interpretadas o manejadas tardía. | Rosero y Ortega (2025) evidencian que la falta de formación incrementa los errores operativos y eleva los costos de mantenimiento. |

*Nota.* Elaboración propia

Los estudios muestran que la adopción de sistemas híbridos, donde se combina inteligencia artificial con gobernanza clínica, presenta mejores indicadores de protección frente a ataques sofisticados, además, se resalta que ningún sistema aislado es suficiente; por ello, los análisis apuntan a modelos integrales de defensa.

**Tabla 7**

*Comparación de Soluciones de Seguridad en Radiología Digital*

| Estrategia | Resultados | Efectividad | Autores |
|---|---|---|---|
| Modelos híbridos IA y supervisión clínica | Reducción significativa de errores diagnósticos al combinar IA con validación médica | Media-Alta | Aguirre et al., (2021) |
| Firmas digitales y trazabilidad (Blockchain) | Garantizan autenticidad e integridad del registro clínico | Alta | Rodríguez y Tafur (2024) |
| Segmentación de redes y control de accesos (Zero-Trust) | Disminuye accesos no autorizados a sistemas PACS y DICOM | Media | Akinci et al., (2025) |
| Validación multimodal (imagen y datos clínicos) | Detecta inconsistencias en diagnósticos generados por IA | Media-Alta | Qureshi y Koo (2026) |
| Autenticación robusta y control humano | Reduce errores operativos y accesos indebidos | Media | Herrera (2024) |
| Monitoreo continúo basado en IA | Identifica patrones anómalos en tiempo real, pero con limitaciones ante ataques avanzados | Media | Poudel et al., (2023) |

*Nota.* Elaboración propia

La ciberseguridad en radiología digital debe evolucionar hacia modelos integrales, adaptativos y multicapa, porque las amenazas GAN y ataques adversariales avanzan más rápido que las soluciones.

Aguirre et al., (2021) subrayan que la IA mejora la eficiencia diagnóstica, pero García et al., (2023) advierten que su uso sin supervisión humana introduce riesgos éticos y clínicos. Por ello, la combinación de IA con validación médica permite reducir errores derivados de manipulaciones, especialmente en contextos donde los deepfakes pueden simular hallazgos patológicos con alto realismo (Aggarwal et al., 2021).

Asimismo, Khalid et al., (2023) subrayan que tecnologías como blockchain, firmas digitales y cifrado avanzado fortalecen la integridad del registro clínico al garantizar la inmutabilidad y trazabilidad de las imágenes. Esto resulta crítico frente a ataques que buscan alterar estudios sin dejar evidencia visible Rodríguez y Tafur (2024).

Igualmente, Akinci et al., (2025) demuestran que una de las principales debilidades es la falta de segmentación de redes hospitalarias y controles de acceso robustos. La implementación de modelos tipo Zero-Trust reduce la superficie de ataque, aunque su adopción es limitada por restricciones tecnológicas y económicas (Herrera, 2024).

Además, Qureshi y Koo (2026) señalan que la integración de múltiples fuentes de información, como imagen, historia clínica, reportes, permite detectar inconsistencias que los modelos de IA no identifican de forma aislada. Esta estrategia es especialmente útil frente a manipulaciones generadas por GAN, que aún presentan limitaciones en coherencia clínica global (Contreras y Pabón, 2025).

Según García et al., (2023) la seguridad no depende únicamente de la tecnología. La capacitación del personal, la ética profesional y los controles institucionales son fundamentales para prevenir accesos indebidos, errores operativos y vulnerabilidades internas. Aunque Poudel et al., (2023) indican que los sistemas de monitoreo detectan comportamientos anómalos, estos presentan limitaciones frente a ataques sofisticados que imitan patrones normales. Esto evidencia que la IA, aunque es parte de la solución, también es parte del problema cuando no se controla adecuadamente (Brohi y Mastoi, 2025).

**Propuesta de Lineamientos, Protocolos de Seguridad y Estrategias Regulatorias, basado en Evidencia**

Con base en la evidencia recopilada, es posible establecer una serie de lineamientos que permitan fortalecer la protección, autenticidad y gobernanza de la IA en entornos sanitarios.

En primer lugar, se propone la implementación de infraestructuras de seguridad híbridas, que integren cifrado avanzado, firmas digitales y sistemas de trazabilidad basados en blockchain, tal como señalan Rodríguez y Tafur (2024) respecto a la necesidad de garantizar una cadena de custodia inalterable, y Pérez (2023) junto con Qureshi y Koo (2026) al destacar la importancia del cifrado robusto contra ataques MitM.

De esta manera, se plantea la adopción de protocolos obligatorios de validación de imágenes médicas, incluyendo watermarking forense para identificar alteraciones no autorizadas, siguiendo las recomendaciones de Pelekis et al. (2025).

En privacidad y gobernanza de datos, los sistemas sanitarios deben incorporar modelos descentralizados como el aprendizaje federado, el cual, según Khalid et al. (2023), minimiza la exposición de información sensible y reduce el riesgo de envenenamiento de datos. Paralelamente, se recomienda generar estándares institucionales de robustez algorítmica mediante el uso de entrenamiento adversarial y auditorías periódicas para evaluar vulnerabilidades, según Zhou et al. (2022) la posibilidad de comprometer la precisión diagnóstica. Brohi y Mastoi (2025) evidencian que los ataques basados en GAN evolucionan más rápido que las defensas actuales, es necesario establecer marcos regulatorios dinámicos, con monitoreo continuo, obligación de reportes de incidentes, certificación periódica de modelos y comités éticos-tecnológicos que supervisen la integridad del ecosistema IA. En conjunto, este

enfoque integral y basada en múltiples capas de defensa permite avanzar hacia una IA más segura, verificable y gobernable dentro de los servicios de salud.

## Conclusiones

La evidencia revisada confirma que las redes generativas antagónicas (GAN) representan una amenaza crítica para la radiología digital, ya que permiten alterar imágenes médicas de manera imperceptible, generando hallazgos falsos que afectan la exactitud diagnóstica, la seguridad del paciente y la confiabilidad institucional. Este tipo de manipulación compromete directamente la integridad del registro clínico y constituye un riesgo ético y operativo para la práctica médica.

De esta forma, los resultados concluyen que gran parte de los sistemas PACS, DICOM y las redes hospitalarias operan con infraestructura obsoleta, configuraciones débiles y ausencia de protocolos de seguridad robustos. Más del 50% de los dispositivos biomédicos continúan trabajando con software o firmwares desactualizados, lo que facilita ataques adversariales, inyección de imágenes falsas e interceptación del flujo de datos clínicos. La falta de estandarización, segmentación de red y auditorías permanentes amplifica estas vulnerabilidades y permite que las manipulaciones ocurran sin trazabilidad ni detección.

Además, las defensas actuales, tecnologías como blockchain, cifrado end-to-end, análisis forense, robust training, IDS/IPS y mecanismos de autenticación aportan mitigaciones parciales, pero ninguna garantiza protección completa. Aunque blockchain alcanza niveles de integridad superiores al 90%, otras estrategias como análisis forense o sistemas IDS/IPS no superan el 63-71% de efectividad, y el robust training reduce entre un 4% y 7% la precisión diagnóstica. Esto evidencia que la velocidad de evolución de los ataques GAN supera la capacidad de respuesta de las defensas disponibles.

Igualmente, se concluye que los desafíos identificados demuestran que la mitigación no es únicamente tecnológica: depende de la infraestructura, del talento humano, de los vacíos

regulatorios y de la gobernanza institucional. La implementación de soluciones avanzadas enfrenta barreras importantes en entornos con recursos limitados, falta de capacitación del personal, ausencia de protocolos de ciberseguridad clínica y marcos normativos que no incluyen lineamientos específicos contra deepfakes médicos o manipulación adversarial.

Igualmente, se concluye que a protección efectiva de la radiología digital requiere una estrategia multicapa que combine la robustez algorítmica, la detección forense avanzada, la seguridad criptográfica y de red, auditorias continuas, gobernanza y regulación actualizada, cadenas de custodia digitales inquebrantables y formación continua del personal.

En síntesis, este enfoque integral es preciso para garantizar la autenticidad de las imágenes médicas, la seguridad del paciente y la confiabilidad del sistema sanitario. Además, las conclusiones permiten afirmar que la manipulación adversarial no es solo un fallo técnico: constituye una amenaza directa contra la seguridad del paciente, vulnera principios bioéticos esenciales como la beneficencia y la no maleficencia, y exige la evolución inmediata de los marcos regulatorios sanitarios hacia modelos más dinámicos, adaptativos y centrados en la protección del diagnóstico médico.

## Referencias Bibliográficas

Abdullah, A., & Vijey, T. (2025). Cybersecurity for analyzing artificial intelligence (AI)-based assistive technology and systems in digital health. Systems, 13(6), 439. https://doi.org/10.3390/systems13060439

Aggarwal, A., Mittal, M., & Battineni, G. (2021). *Generative adversarial network: An overview of theory and applications*. *International Journal of Information Management Data Insights, 1*(2), 100004. https://doi.org/10.1016/j.jjimei.2020.100004

Aguilar, J. (2025). *Impacto y degradación de modelos explicables de machine learning aplicados a las áreas de salud, seguridad y defensa nacional de España ante ataques de envenenamiento de datos* [Trabajo de fin de máster, Universidad Europea]. https://titula.universidadeuropea.com/bitstream/handle/20.500.12880/14088/TFM_Jenny_Aguilar-Guinez.pdf?sequence=1&isAllowed=y

Aguirre, F., Carballo, L., González, X., & Gigirey, V. (2021). *Inteligencia artificial aplicada a la imagen médica*. Revista de Imagenología. https://www.sriuy.org.uy/ojs/index.php/Rdi/article/view/94

Akinci, T., Tejani, A., Khosravi, B., Bluethgen, C., Busch, F., Bressem, K., Adams, L., Moassefi, M., Faghani, S., & Gichoya, J. (2025). Cybersecurity threats and mitigation strategies for large language models in health care. Radiology: Artificial Intelligence. https://doi.org/10.1148/ryai.240739

Ayerbe, A. (2020). *La ciberseguridad y su relación con la inteligencia artificial*. Real Instituto Elcano. https://media.realinstitutoelcano.org/wp-content/uploads/2021/10/ari128-2020-ayerbe-ciberseguridad-y-su-relacion-con-inteligencia-artificial.pdf

Biasin, E., Kamenjašević, E., & Ludvigsen, K. (2024). Cybersecurity of AI medical devices: Risks, legislation, and challenges. En Law 2024 (pp. 57–74). Edward Elgar Publishing. https://doi.org/10.4337/9781802205657.ch04

Bortsova, G., González, C., Wetstein, S., Dubost, F., Katramados, I., Hogeweg, L., Sánchez, C. (2021). Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. Medical Image Analysis, 73, Article 102141. https://doi.org/10.1016/j.media.2021.102141

Brohi, S., & Mastoi, Q. (2025). AI Under Attack: Metric-Driven Analysis of Cybersecurity Threats in Deep Learning Models for Healthcare Applications. Applied Sciences, 18(3), 157. https://doi.org/10.3390/a18030157

Contreras, E., & Pabón, H. (2025). *Inteligencia artificial como complemento de apoyo diagnóstico, revisión a partir de una pasantía académica en España*. Repositorio UDES. https://repositorio.udes.edu.co/server/api/core/bitstreams/db112532-6109-4c06-8413-e5186b73f7e0/content

Eichelberg, M., Kleber, K., & Kämmerer, M. (2020). Cybersecurity challenges for PACS and medical imaging. Academic Radiology, 27(8), 1126–1139. https://www.academicradiology.org/article/S1076-6332(20)30171-9/fulltext

Elsevier. (s. f.). *Medical image data*. ScienceDirect Topics. https://www.sciencedirect.com/topics/computer-science/medical-image-data

García, A., Girón, F., & Rosselli, D. (2023). La integración de la inteligencia artificial en la atención médica: Desafíos éticos y de implementación. *Universitas Médica, 64*(3). https://doi.org/10.11144/Javeriana.umed64-3.inte

George Lawton, *Healthcare Has Mixed Feelings about AI*, IEEE INTELLIGENT SYSTEMS (May/June 2012), http://www.computer.org/intelligent.

Hernández, R., Fernández, C. & Baptista, P. (2014). Metodología de la Investigación. Edición 6. McGraw Hill. https://dialnet.unirioja.es/servlet/libro?codigo=775008

Herrera, C. (2024). *Modelo de seguridad informática para la evaluación de dispositivos biomédicos*. Repositorio Universidad Distrital Francisco José de Caldas. https://repository.udistrital.edu.co/server/api/core/bitstreams/c40639e6-5d09-44fd-88d3-d26efd4ed0f6/content

IT Masters Magazine. (2024). *Envenenamiento de datos: Cómo proteger la IA en un mundo de riesgos reales*. https://www.itmastersmag.com/ciberseguridad/el-envenenamiento-de-datos-un-peligro-para-toda-la-inteligencia-artificial/

Johnson, S. L. J. (2019). AI, Machine Learning, and Ethics in Health Care. Journal of Legal Medicine, 39(4), 427–441. https://doi.org/10.1080/01947648.2019.1690604

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. Computers in Biology and Medicine, 163, 106848. https://doi.org/10.1016/j.compbiomed.2023.106848

Minagi, A., Hirano, H., & Takemoto, K. (2022). Natural images allow universal adversarial attacks on medical image classification using deep neural networks with transfer learning. Journal of Imaging, 8(2), 38. https://doi.org/10.3390/jimaging8020038

Open MedScience. (2025, August 10). The role of cybersecurity in medical imaging systems. https://openmedscience.com/the-role-of-cybersecurity-in-medical-imaging-systems/

Pantoja, C., Mosquera, E., Delgado, K., López, K., & Gómez, L. (2025). *Optimización automática de parámetros de adquisición en imágenes diagnósticas mediante algoritmos*

*de IA y aprendizaje automático* [Trabajo de grado, Universidad Nacional Abierta y a Distancia]. Repositorio Institucional UNAD. https://repository.unad.edu.co/jspui/bitstream/10596/78435/1/Lvgomezpr.pdf

Pelekis, S., Koutroubas, T., Blika, A., Berdelis, A., Karakolis, E., Ntanos, C., Spiliotis, E., & Askounis, D. (2025). Adversarial machine learning: A review of methods, tools, and critical industry sectors. *Artificial Intelligence Review* (*58),* 226. https://doi.org/10.1007/s10462-025-11147-4

Pérez, M. (2023). *Seguridad de los dispositivos médicos: Lo que deben saber los fabricantes de equipos originales*. Digi International. https://es.digi.com/blog/post/medical-device-security

Poudel, R., Rahman, M., Rahman, M., & Rahman, M. (2023). Adversarial attacks on AI systems: A growing cyber threat. *International Journal of Science and Research Archive, 10*(2), 1086–1092. https://doi.org/10.30574/ijsra.2023.10.2.1086

Qureshi, R., & Koo, I. (2026). A comprehensive survey of cybersecurity threats and data privacy issues in healthcare systems. Applied Sciences, 16(3), 1511. https://doi.org/10.3390/app16031511

Rankin, B. (2024). AI, adversarial attacks, and insider threats in life sciences. USDM Life Sciences. https://usdm.com/resources/blogs/adversarial-attacks-and-insider-threats-in-life-sciences

Rodríguez, A., & Tafur, B. (2024). *Diseño de un modelo de gestión para la protección de datos personales en la prestación de servicios de salud con inteligencia artificial en Colombia* [Trabajo de grado]. Politécnico Grancolombiano.

https://alejandria.poligran.edu.co/bitstream/handle/10823/7824/INFORME%203%20-%20TRABAJO%20DE%20GRADO%20MGP%20final%20conclusiones.pdf

Sullivan, D. (2015, November 4). *How machine learning works, as explained by Google*. MarTech. https://martech.org/how-machine-learning-works/

Zhou, S., Liu, C., Ye, D., Zhu, T., Zhou, W., & Yu, P. (2022). Adversarial attacks and defenses in deep learning: From a perspective of cybersecurity. *ACM Computing Surveys, 55*(8). https://doi.org/10.1145/3547330