PUBLICACIONES DE LA
ACADEMIA NACIONAL DE
MEDICINA DE MÉXICO

ACTUALIDADES EN INTELIGENCIA ARTIFICIAL

Dr. Rodolfo Palencia Díaz
Dr. Rodolfo de J. Palencia Vizcarra
Dr. Raúl Carrillo Esper

Número 4

Inteligencia Artificial y Educación Médica Continua: Implicaciones y Ruta Estratégica para la Academia Nacional de Medicina de México.

Dr. Rodolfo Palencia Díaz
Dr. Rodolfo de J. Palencia Vizcarra

Médicos Internistas
Universidad de Guadalajara
Instituto Mexicano del Seguro Social (IMSS)
Colegiados y Certificados (CMMI)
Fundadores del TICC
10 de marzo 2026

Introducción

La inteligencia artificial (IA), con énfasis en la IA generativa, ha dejado de ser una innovación periférica para convertirse en un determinante operativo fundamental en la formación médica actual. Para la Academia Nacional de Medicina de México, la integración de estas herramientas no es solo una opción tecnológica, sino una necesidad estratégica para redefinir cómo los médicos aprenden, actualizan sus conocimientos y validan su práctica profesional.

El volumen de información médica crece a una velocidad que supera la capacidad de procesamiento humano, exigiendo una transición hacia modelos
de "aprendizaje aumentado". En este esquema, la IA funciona como un tutor adaptativo y un apoyo para el razonamiento clínico, mientras que el médico conserva la responsabilidad final de la contextualización y la toma de decisiones.

Análisis de la Evidencia Reciente

La literatura científica actual presenta un panorama dual: un alto potencial de optimización educativa frente a una base de evidencia que aún es metodológicamente inmadura.

Hallazgos Clave de Revisiones Sistemáticas y Guías BEME.

De acuerdo con la BEME Guide No. 84 y diversas revisiones sistemáticas de 2024 y 2025, se han identificado las siguientes aplicaciones y realidades:

- Alcance: La IA ya permea todo el continuo formativo (pregrado, posgrado y educación médica continua o EMC), con usos documentados en admisión, enseñanza, evaluación y razonamiento clínico.
- Eficacia: Se observa una mejora en habilidades prácticas y diagnóstico. Sin embargo, los estudios actuales suelen ser unicéntricos y con muestras pequeñas, lo que limita la generalización de los resultados.
- Riesgos Cognitivos: Existe un riesgo latente de empobrecimiento del aprendizaje si se sustituye el esfuerzo cognitivo del médico por la producción automática de la IA, así como preocupaciones críticas sobre la exactitud del contenido y la integridad académica.

Síntesis Comparativa de la Evidencia

Fuente	Tipo de Estudio	Aporte Principal para la EMC	Limitación Identificada
Gordon et al., 2024 (BEME 84)	Scoping Review	Mapeo completo de usos en enseñanza y razonamiento clínico.	Alta heterogeneidad; sin efecto educativo uniforme.
Tozsin et al., 2024	Revisión Sistemática	Mejora documentada en habilidades prácticas y evaluación.	Validación insuficiente y evidencia desigual.
Feigerlova et al., 2025	Revisión Sistemática	Medición de desenlaces educativos específicos.	Debilidad metodológica; estudios unicéntricos.
Kovalainen et al., 2025	Umbrella Review	Identificación de dominios (Robótica, NLP, Big Data).	No exclusivo para médicos o EMC.
Masters, 2023/2025 (AMEE 158/178)	Guías AMEE	Estructuración de ética, evaluación y desarrollo docente.	No miden efectividad; son guías de implementación.

Implicaciones para la Educación Médica Continua (EMC)

La implementación de la IA en la EMC requiere un desplazamiento de la conversación desde el simple "uso de la herramienta" hacia un "uso con legitimidad ética".

1. Cambio Pedagógico y Evaluativo

La IA no debe ser un sustituto del estudio. La evidencia favorece su uso como:

- Simulador y generador de casos clínicos complejos.
- Sintetizador preliminar de literatura científica.
- Apoyo para retroalimentación inmediata.

En cuanto a la evaluación, el auge de modelos generativos vuelve obsoletos los instrumentos tradicionales (como tareas escritas no supervisadas). Se propone un rediseño basado en:

- Auditoría de prompts y salidas de la IA.
- Defensa oral del razonamiento clínico.
- Tareas de detección de sesgos, errores o "alucinaciones" del sistema.

2. Marco de Competencias Digitales (Consenso DECODE 2025)

Para que la alfabetización en IA sea efectiva, debe integrarse como una competencia transversal en cuatro dominios principales:

1. Profesionalismo en salud digital.
2. Salud digital del paciente y de la población.
3. Sistemas de información en salud.
4. Ciencia de datos en salud.

Ruta de Implementación para la Academia Nacional de Medicina

La prioridad institucional debe ser el establecimiento de estándares de calidad y seguridad, más que la recomendación de plataformas específicas. Se propone una estructura en cuatro niveles progresivos:

1. Alfabetización Básica: Comprender alcances, límites y tipos de IA.
2. Uso Seguro en Docencia y Actualización: Aplicación práctica en el entorno clínico-educativo.
3. Evaluación Crítica de Herramientas: Capacidad para auditar la evidencia generada por algoritmos.
4. Implementación con Gobernanza: Aplicación de principios institucionales y éticos.

Matriz Mínima de Seguridad para el Uso de IA

Cualquier programa formativo que incorpore IA debe definir:

- Finalidad del uso: ¿Para qué se emplea la herramienta?
- Tipo de dato: Protección y anonimización de información clínica.
- Supervisión humana: El juicio médico como filtro final.
- Fuente de verificación: Contraste obligatorio con guías y fuentes primarias.

Gobernanza Ética y Estándares Internacionales

La adopción de IA debe regirse por los principios del consenso FUTURE-AI, los cuales son transferibles a la educación médica continua:

- Fairness (Equidad): Mitigación de sesgos.
- Universality (Universalidad): Aplicabilidad en diversos contextos.
- Traceability (Trazabilidad): Seguimiento de cómo se generó la información.
- Usability (Usabilidad): Facilidad de integración en el flujo de trabajo.
- Robustness (Robustez): Fiabilidad técnica del sistema.
- Explainability (Explicabilidad): Claridad sobre el razonamiento algorítmico.

Asimismo, el estándar GAMER Statement (2025) debe ser la referencia para reportar el uso de IA generativa en cualquier investigación o material educativo producido bajo el aval de la Academia.

Conclusiones

La IA incrementa la necesidad de las habilidades clínicas clásicas. La lectura crítica, el análisis de sesgos y la metacognición son hoy más vitales que nunca para actuar como contrapesos del asistente algorítmico.
Para los miembros de la Academia Nacional de Medicina de México, el imperativo es claro: la meta no es crear dependencia tecnológica, sino formar un clínico capaz de aprovechar la velocidad de la IA sin sacrificar el rigor científico ni la ética profesional. La IA no reduce la necesidad de juicio experto; por el contrario, la eleva a un nivel de supervisión y auditoría indispensable para la seguridad del paciente.

Referencias Bibliográficas de Impacto

1. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide No.

84. Med Teach. 2024;46(4):446-70. doi:10.1080/0142159X.2024.2314198.
2. Tozsin A, Ucmak H, Soyturk S, Aydin A, Gozen AS, Al Fahim M, et al. The role of artificial intelligence in medical education: a systematic review. Surg Innov. 2024;31(4):415-23. doi:10.1177/15533506241248239.
3. Car J, Lindberg O, Azzopardi-Muscat N, et al. The Digital Health Competencies in Medical Education Framework: an international consensus statement based on a Delphi study. JAMA Netw Open. 2025;8(1):e2453131.doi:10.1001/jamanetworkopen.2024.53131.
4. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. Med Teach. 2023;45(6):574-84. doi:10.1080/0142159X.2023.2186203.
5. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:e081554. doi:10.1136/bmj-2024-081554.
6. Luo X, O'Connor DB, Liu N, et al. Reporting guideline for the use of Generative Artificial intelligence tools in MEdical Research: the GAMER Statement. BMJ Evid Based Med. 2025;30(6):390-400. doi:10.1136/bmjebm-2025-113825

Inteligencia artificial y educación médica continua

Implicaciones, evidencia reciente y ruta de implementación para la Academia Nacional de Medicina de México

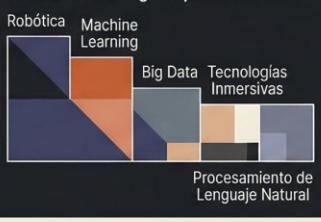
IA EN EDUCACIÓN MÉDICA CONTINUA: HACIA EL APRENDIZAJE AUMENTADO

La inteligencia artificial (IA) ha pasado de ser periférica a ser un **determinante operativo** de la formación médica. Esta revisión propone una ruta para la Academia, transitando del simple uso de herramientas hacia un modelo de "aprendizaje aumentado" con gobernanza ética.

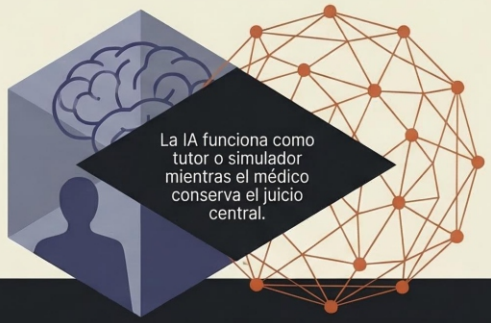
EVIDENCIA Y ESTADO ACTUAL

Potencial alto con evidencia inmadura.

Dominios tecnológicos predominantes



EL MODELO DE APRENDIZAJE AUMENTADO



IMPLEMENTACIÓN Y COMPETENCIAS

LOS CUATRO NIVELES DE LA ACADEMIA

REDISEÑO DE LA EVALUACIÓN MÉDICA

Transitar hacia casos contextualizados y auditoría de "prompts" para evitar la dependencia cognitiva.

PRINCIPIOS FUTURE-AI PARA GOBERNANZA

DOMINIOS COMPETENCIALES ESENCIALES (MARCO DECODE/AMEE)

Dominio	Competencia Esencial
Alfabetización	Comprender tipos de IA, sus alcances técnicos y límites operativos.
Ética y Legal	Proteger datos clínicos, declarar uso de IA y reconocer sesgos algorítmicos.
Integración	Usar la IA como apoyo crítico sin sustituir el razonamiento clínico final.

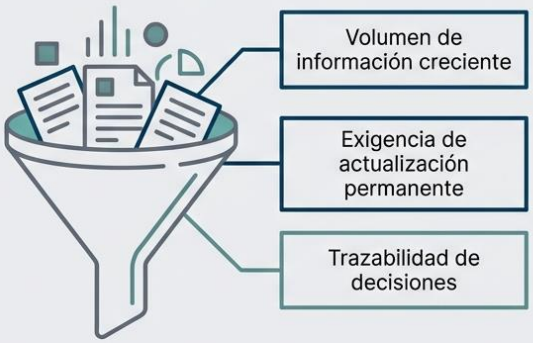
REFERENCIAS BIBLIOGRÁFICAS (FORMATO VANCOUVER)

1. Gordon M et al. Med Teach. 2024;48(4):445-70. 2. Gar J et al. JAMA Netw Open. 2025;8(1):e2483131. 3. Leliadir K et al. BMJ. 2023;988:e061554. 4. Mosters K. Med Teach. 2023;45(8):574-64. S. Luo X et al. BMJ Evid Based Med. 2025,30(8):390-400.

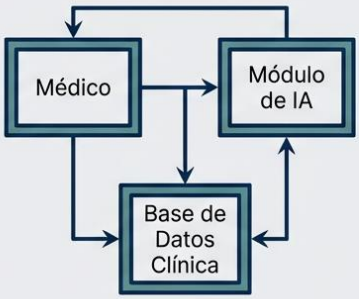
NotebookLM

El Motivo de Consulta: Presión Inédita sobre la EMC

Sobrecarga Cognitiva



Determinante Operativo



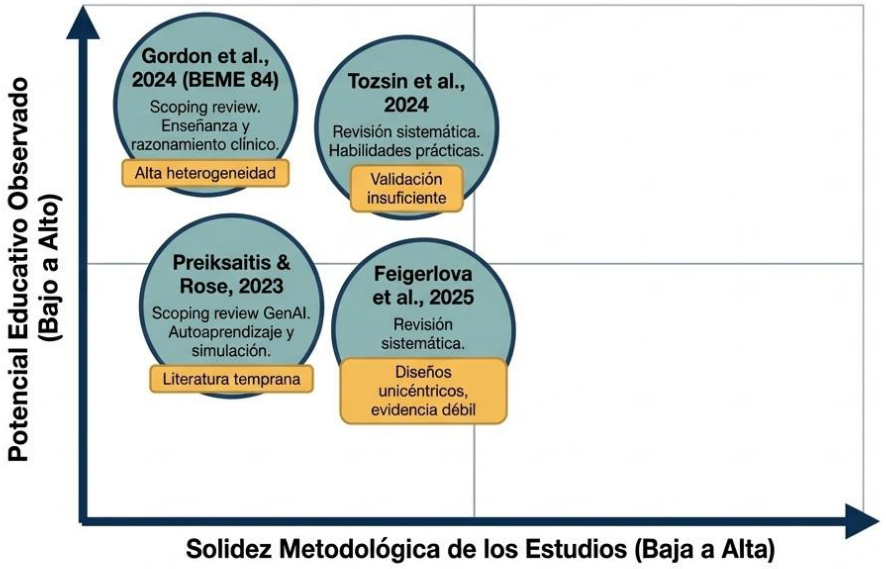
La IA modifica simultáneamente el contenido a aprender, las herramientas de aprendizaje y los criterios de validación.

"La alfabetización en IA es una competencia transversal de la práctica médica contemporánea, no una curiosidad tecnológica."

8. Kovalainen T, et al. Nurse Educ Today. 2025; 9. Car J, et al. JAMA Netw Open. 2025; 10. Masters K. Med Teach. 2023; 11. Masters K, et al. Med Teach. 2025; 12. Tolsgaard MG, et al. Med Teach. 2

NotebookLM

El Panorama Dual de la Evidencia Actual



Conclusión Operativa

La pregunta clínica ya no es si la IA 'sirve', sino bajo qué condiciones institucionales y pedagógicas genera valor sin degradar el aprendizaje.

4. Gordon M, et al. Med Teach. 2024; 5. Tozsin A, et al. Surg Innov. 2024; 6. Preiksaitis C, Rose C. JMIR Med Educ. 2023; 7. Feigerlova E, et al. BMC Med Educ. 2025.

NotebookLM

Anatomía de la Educación Médica en la Era de la IA

8. Kovalainen T, et al. Nurse Educ Today. 2025; 16. Lekadir K, et al. BMJ. 2025.

NotebookLM

Implicación Pedagógica: Del Reemplazo al Aumento

5. Tozsin A, et al. Surg Innov. 2024; 6. Preiksaitis C, Rose C. JMIR Med Educ. 2023; 7. Feigerlova E, et al. BMC Med Educ. 2025; 11. Masters K, et al. Med Teach. 2025; 12. Tolsgaard MG, et al. Med T

NotebookLM

Implicación Evaluativa: Rediseño de Instrumentos de EMC

EMC Tradicional (Vulnerable a GenAI)	EMC Robusta (Adaptada a la IA)
✕ Tareas escritas no supervisadas.	✓ Evaluación basada en casos clínicos contextualizados.
− Pruebas de baja autenticidad y memorización.	✓ Auditoría de "prompts" (instrucciones) y verificación de fuentes.
− Enfoque en la respuesta final.	✓ Defensa oral de razonamiento.
	✓ Tareas de **detección de errores, sesgos** o **"alucinaciones"** del sistema.

La respuesta institucional no es el prohibicionismo, sino el rediseño auténtico.

6. Preiksaitis C, Rose C. JMIR Med Educ. 2023; 11. Masters K, et al. Med Teach. 2025.

NotebookLM

Gobernanza: Implicaciones Éticas y Organizacionales

Ética y Regulación

Reglas explícitas obligatorias:

- Anonimización de casos reales
- Propiedad de la información
- Trazabilidad y supervisión humana

únicamente *"cómo usar"* herramientas, sino *"cuándo no usarlas"*, *"qué verif siempre"* y *"qué grado de confianza merecen según la tarea"*. La guía ética AMEE y

Estrategia Organizacional

Mecanismos de adopción:

- Formatos flexibles y asincrónicos.
- Microcredenciales y talleres de casos clínicos con IA.
- Necesidad crítica de formación docente y apoyo institucional.

10. Masters K. Med Teach. 2023; 11. Masters K, et al. Med Teach. 2025; 16. Lekadir K, et al. BMJ. 2025; 17. Henning PA, et al. J Eur CME. 2021.

NotebookLM

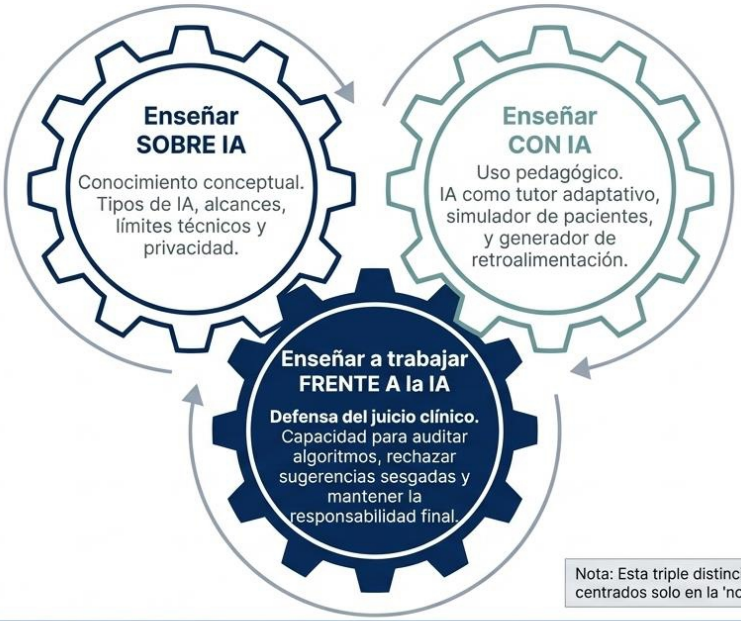
El Fulcro Epistemológico: El Contrapeso Indispensable

> *"En términos prácticos, la IA no reduce la necesidad de juicio experto en la práctica médica; la incrementa."*

6. Preiksaitis C, Rose C. JMIR Med Educ. 2023; 7. Feigerlova E, et al. BMC Med Educ. 2025; 10. Masters K. Med Teach. 2023; 12. Tolsgaard MG, et al. Med Teach. 2023.

NotebookLM

Los Tres Planos del Aprendizaje en IA



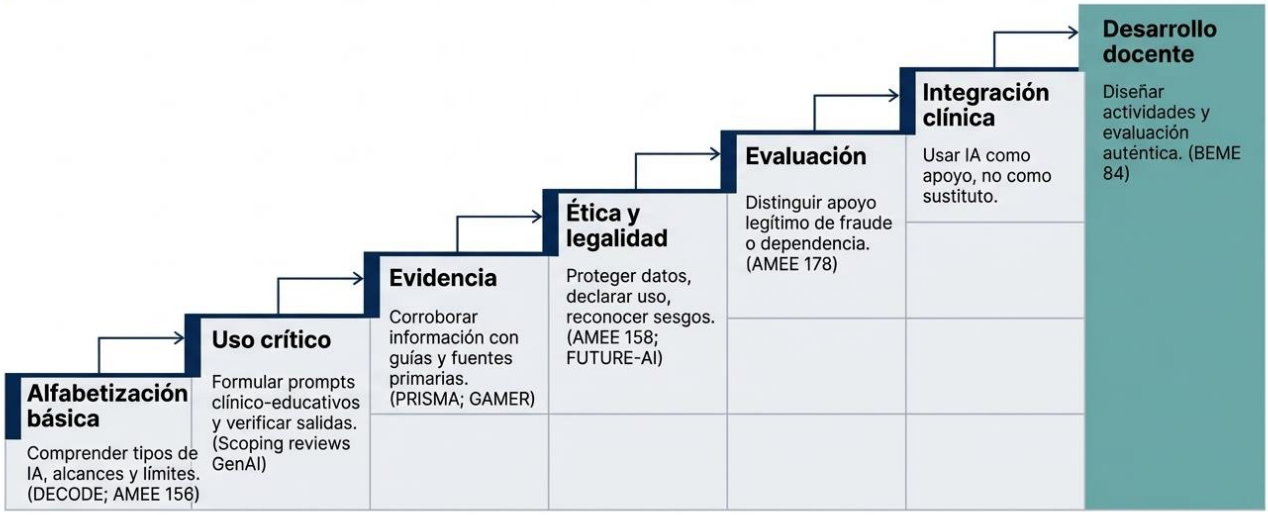
Nota: Esta triple distinción evita programas superficiales centrados solo en la 'novedad tecnológica'.

4. Gordon M, et al. *Med Teach*. 2024; 6. Preiksaitis C, Rose C. *JMIR Med Educ*. 2023; 9. Car J, et al. *JAMA Netw Open*. 2025; 11. Masters K, et al. *Med Teach*. 2025.

NotebookLM

Academic Clinical Dashboard

Matriz de Competencias Mínimas para el Médico Internista



3. Luo X, et al. BMJ Evid Based Med. 2025; 9. Car J, et al. JAMA Netw Open. 2025; 10. Masters K. Med Teach. 2023; 11. Masters K, et al. Med Teach. 2025; 12. Tolsgaard MG, et al. Med Teach. 2023; 16. Lekadir K, et al. BMJ. 2025.

NotebookLM

Ruta de Implementación para la Academia Nacional de Medicina de México

Fase 1: Alfabetización Básica Estructurada

Desarrollo de microcredenciales. Inserción en programas de profesionalismo y gestión de información.

Fase 2: Uso Seguro en Docencia y Actualización

Integración de IA en casos mexicanos, escenarios de urgencias, consultorio y hospital.

Fase 3: Evaluación Crítica de Herramientas

Matriz mínima de seguridad: finalidad, tipo de dato, supervisión humana, y "criterios de suspensión".

Fase 4: Gobernanza Institucional

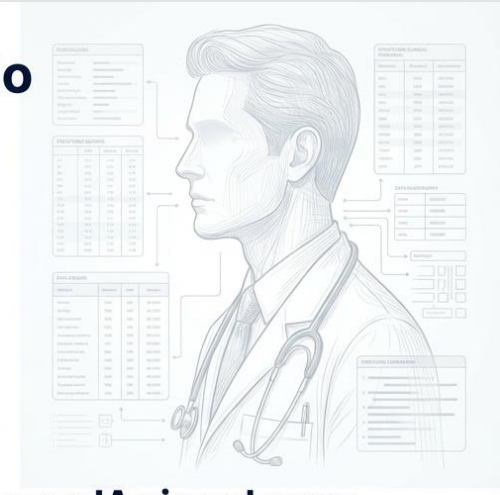
Implementación de lineamientos FUTURE-AI y GAMER. Establecimiento de estándares, no recomendación de plataformas comerciales únicas.

conversación desde "usar la herramienta" hacia "usar la herramienta con legitimidad"

9. Car J, et al. JAMA Netw Open. 2025; 10. Masters K. Med Teach. 2023; 11. Masters K, et al. Med Teach. 2025; 12. Tolsgaard MG, et al. Med Teach. 2023; 16. Lekadir K, et al. BMJ. 2025.

NotebookLM

Conclusión: El Clínico Aumentado frente al Algoritmo



- La postura institucional no debe ser el entusiasmo acrítico ni el rechazo defensivo, sino una adopción gradual, gobernada, evaluable y centrada en competencias.
- El médico que no interactúe críticamente con IA quedará rezagado; el que delegue sin criterio, también.

> **“La meta es formar a un clínico capaz de usar IA sin volverse dependiente, de aprovechar su velocidad sin sacrificar rigor, y de mantener el razonamiento médico, la ética y la responsabilidad final como núcleos irrenunciables de la práctica.”**

4. Gordon M, et al. Med Teach. 2024; 5. Tozsin A, et al. Surg Innov. 2024; 6. Preiksaitis C, Rose C. JMIR Med Educ. 2023; 7. Feigerlova E, et al. BMC Med Educ. 2025; 8. Kovalainen T, et al. Nurse Educ Today. 2025; 9. Car J, et al. JAMA Netw Open. 2025; 10. Masters K. Med Teach. 2023; 11. Masters K, et al. Med Teach. 2025; 12. Tolsgaard MG, et al. Med Teach. 2023; 16. Lekadir K, et al. BMJ. 2025.

Detección y Mitigación de Sesgos y Alucinaciones en Inteligencia Artificial Generativa: Una Revisión Crítica para la Práctica Clínica y la Investigación

Dr. Rodolfo Palencia Díaz
Dr. Rodolfo de J. Palencia Vizcarra

Médicos Internistas
Universidad de Guadalajara
Instituto Mexicano del Seguro Social (IMSS)
Colegiados y Certificados (CMMI)
Fundadores del TICC
11 de marzo 2026

La integración de modelos de lenguaje de gran tamaño (LLM) y otras formas de inteligencia artificial (IA) en la medicina contemporánea ha generado una paradoja práctica. Si bien estas herramientas optimizan la velocidad de acceso y síntesis de la información, también introducen errores plausibles, sesgos distributivos y respuestas fabricadas con una apariencia de autoridad técnica. Para el médico internista y el investigador, la seguridad clínica hoy depende menos de la sofisticación tecnológica del modelo y más de la calidad del flujo humano-IA y de la capacidad del usuario para interrogar y contrastar los resultados.

Definiciones Operativas y la Lógica de Verificación

Para un uso responsable de la IA, es imperativo abandonar la idea de que la tecnología "sabe" y reemplazarla por una lógica de verificación. La salida del modelo debe ser tratada siempre como una hipótesis inicial, no como evidencia final. En términos operativos, se definen dos fenómenos críticos:

- Alucinación: Generación de contenido inexacto, no sustentado o inventado que suele presentarse con un tono convincente y fluido.
- Sesgo: Desviación sistemática que perjudica la exactitud, la justicia o la aplicabilidad de la respuesta en determinados grupos, escenarios clínicos o contextos de uso.

Dominios de Error y Riesgo Clínico

La evidencia científica permite categorizar los problemas de la IA generativa en dos dominios mayores: la alucinación factual y el sesgo sistemático. El peligro radica en que las salidas incorrectas suelen ser internamente coherentes y persuasivas, lo que aumenta el riesgo de una automatización acrítica, especialmente en contextos de alta carga asistencial o fatiga clínica.

Tabla 1. Reconocimiento de Sesgos y Alucinaciones en la Práctica Clínica

Hallazgo en la Salida del Modelo	Interpretación Probable	Riesgo Clínico
Referencias con títulos plausibles pero imposibles de localizar.	Alucinación bibliográfica.	Sustentar decisiones en evidencia inexistente.
Recomendaciones idénticas en redacción, pero distintas por sexo, etnia o nivel socioeconómico sin justificación.	Sesgo demográfico.	Inequidad diagnóstica o terapéutica.
Respuesta muy segura ante una pregunta ambigua o con datos insuficientes.	Sobreconfianza algorítmica.	Cierre prematuro del razonamiento clínico.
Resumen "bonito" que omite población, comparador, desenla ces o limitaciones.	Compresión engañosa de evidencia.	Sobreinterpretación de estudios.
Falta de reconocimiento de incertidumbre, contraejemplos o escenarios alternos.	Razonamiento frágil.	Error de priorización y falsa certeza.

Desempeño y Limitaciones de los Modelos

Revisiones sistemáticas y metaanálisis han evaluado la capacidad de la IA en tareas de salud. Aunque demuestran utilidad en educación médica, apoyo documental y comunicación, su capacidad clínica es parcial:

1. Exactitud en Exámenes: En exámenes de salud globales, se ha documentado una exactitud de 0.61, mientras que en el USMLE esta cifra desciende a 0.51.
2. Falla de Generalización: Un modelo puede rendir adecuadamente en preguntas cerradas, pero fallar en el razonamiento contextual, la integración longitudinal de datos del paciente y el manejo de excepciones.
3. Trazabilidad: La bibliografía generada por LLM presenta tasas relevantes de citas inventadas, por lo que no puede aceptarse sin comprobación directa de DOIs o registros editoriales.

Estrategias de Mitigación y Arquitectura de Seguridad

Evitar errores no depende de una sola técnica, sino de una arquitectura de seguridad por capas. La IA debe actuar como un asistente de borrador cognitivo, nunca como un sustituto del juicio clínico.

Tabla 2. Estrategias para Evitar Sesgos y Alucinaciones

Estrategia	Qué Resuelve	Aplicación Práctica para el Médico
Prompt Estructurado	Reduce ambigüedad y deriva semántica.	Especificar objetivo, contexto, población y formato (ej. "separa hechos de dudas").
Verificación de Fuentes	Disminuye la alucinación bibliográfica.	Rechazar referencias que no pueden localizarse físicamente.
RAG (Generación Aumentada por Recuperación)	Reduce errores por conocimiento desactualizado.	Conectar el modelo a un corpus institucional o literatura curada.
Doble Lectura Humana	Reduce la automatización acrítica.	El clínico conserva la autoridad epistémica; la IA solo hace el borrador.
Auditoría por Subgrupos	Detecta sesgo latente.	Probar el mismo caso clínico variando sexo, edad o etnia para observar discrepancias.
Registro de Errores	Mejora la gobernanza local.	Crear una bitácora de fallos en el servicio o comité.

Marcos de Calidad y Reporte en Investigación

Para asegurar la transparencia y la trazabilidad en la era de la medicina digital, es fundamental adherirse a marcos metodológicos robustos. El uso de la IA en la investigación debe ser reportado de forma auditable.

Tabla 3. Marcos de Calidad para el Uso Responsable de IA

Marco	Utilidad Principal	Relevancia para Sesgo/Alucinación
PRISMA 2020	Reporte de revisiones sistemáticas.	Transparencia en la selección y síntesis de evidencia.
AMSTAR-2	Evaluación crítica de revisiones.	Permite juzgar la robustez de la evidencia secundaria.
SPIRIT-AI	Protocolos de ensayos con IA.	Obliga a describir entradas, errores y contexto de interacción.
CONSORT-AI	Reporte de ensayos con IA.	Exige claridad en fallos e integración clínica.
GAMER	Reporte de IA generativa en investigación.	Hace visible dónde y cómo intervino la IA en la concepción o análisis.

Conclusiones y Reglas Prácticas para el Clínico

La IA generativa no posee autoridad autónoma en decisiones complejas. Su valor reside en su capacidad de responder con trazabilidad y justicia, siempre bajosupervisión humana final. Todo médico debe someter las respuestas de la IA a tres filtros fundamentales:

1. Verificabilidad: ¿Puedo rastrear la fuente original?
2. Equidad: ¿Cambiaría la respuesta si modifico un atributo demográfico irrelevante?
3. Plausibilidad Clínica: ¿La salida resiste la confrontación con guías de práctica clínica, la fisiopatología y el contexto específico del paciente?

Referencias Bibliográficas

1. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel). 2023;11(6):887. doi:10.3390/healthcare11060887.
2. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). NPJ Digit Med. 2024;7(1):183. doi:10.1038/s41746-024-01157-x.
3. Omar M, Sorin V, Agbareia R, Apakama DU, Benjamini M, Shilo S, et al. Evaluating and addressing demographic disparities in medical large language models: a systematic review. Int J Equity Health. 2025;24(1):57. doi:10.1186/s12939-025-02419-0.
4. Waldock WJ, Zhang J, Guni A, Nabeel A, Darzi A, Ashrafian H. The accuracy and capability of artificial intelligence solutions in health care examinations and certificates: systematic review and meta-analysis. J Med Internet Res. 2024;26:e56532. doi:10.2196/56532.
5. Fareed M, Fatima M, Uddin J, Ahmed A, Sattar MA. A systematic review of ethical considerations of large language models in healthcare and medicine. Front Digit Health. 2025;7:1653631. doi:10.3389/fdgth.2025.1653631.
6. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: WHO; 2021. ISBN:9789240029200.
7. Chelli M, Gaitanaros S, Pellegrino A, Kumar N, Bredella MA, Fabbri N, et al. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: comparative analysis. J Med Internet Res. 2024;26:e53164. doi:10.2196/53164.
8. Pfohl SR, Yao S, Foryciarz A, Kathuria A, Hegarty-Craver M, Conrad F, et al. A toolbox for surfacing health equity harms and biases in large language models. Nat Med. 2024. doi:10.1038/s41591-024-03297-7.
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

doi:10.1136/bmj.n71.
10. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomized studies of healthcare interventions, or both. BMJ.2017;358:j4008. doi:10.1136/bmj.j4008.
11. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reportsfor interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020;2(10):e537-e548. doi:10.1016/S2589-7500(20)30218-1.
12. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;2(10):e549-e560. doi:10.1016/S2589-7500(20)30219-3.
13. Luo X, Tham YC, Giuffrè M, Ranisch R, Daher M, Lam K, et al. Reporting guideline for the use of Generative Artificial intelligence tools in Medical Research: the GAMER Statement. BMJ Evid Based Med. 2025;30(6):390-400. doi:10.1136/bmjebm-2025-113825.
14. Hake J, Crowley M, Plotkin A, Bell S, Patt D, Chien A, et al. Quality, accuracy, and bias in ChatGPT-based summarization of medical abstracts. Oncologist. 2024. PMID: 38527823.
15. Amugongo LM, Mascheroni P, Brooks S, Doering S, Seidel J. Retrieval augmented generation for large language models in healthcare: a systematic review. PLOS Digit Health. 2025;4(6):e0000877. doi:10.1371/journal.pdig.0000877.

Uso Seguro de la IA Generativa en Medicina

Diagnóstico preventivo de sesgos y alucinaciones para la práctica clínica, docente y de investigación.

Author Profile

Dr. Rodolfo Palencia Díaz.
Dr. Rodolfo de J Palencia Vizcarra.
Médicos Internistas

Universidad de Guadalajara | Instituto Mexicano del Seguro Social IMSS | Colegiados CMIM y CMMI | Fundadores de TICC

NotebookLM

IA Generativa en Medicina: De la Autoridad Algorítmica a la Verificación Clínica

Dr. Rodolfo Palencia Díaz / Dr. Rodolfo de J. Palencia Vizcarra
(Universidad de Guadalajara, IMSS)

La integración de modelos de lenguaje (LLM) en medicina ha creado una paradoja: mayor velocidad de síntesis frente al riesgo de errores plausibles. Esta guía propone abandonar la idea de que la IA "sabe" y reemplazarla por una lógica de **verificación humana rigurosa**.

Identificación del Riesgo: Alucinaciones y Sesgos

Alucinación Factual y Bibliográfica

Generación de contenido inventado, datos inexistentes o referencias bibliográficas falsas con tono convincente.

Sesgo Sistemático y Demográfico

Desviaciones que perjudican la equidad según sexo, raza, edad o nivel socioeconómico.

El Peligro de la Fluidez Lingüística

Las respuestas incorrectas suelen ser coherentes y persuasivas, aumentando el riesgo de automatización acrítica.

Tabla: Diferenciación del Riesgo Clínico

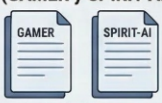
Hallazgo en la IA	Interpretación Probable	Riesgo Clínico
Citas imposibles de localizar	Alucinación bibliográfica	Decisiones basadas en evidencia inexistente
Respuestas idénticas ein ajuste demográfico	Sesgo demográfico	Inequidad diagnóstica o terapéutica
Respuesta segura ante datos insuficientes	Sobreconfianza algorítmica	Cierre prematuro del razonamiento clínico

Estrategias de Mitigación y Marcos de Calidad

El Triple Filtro de Verificación

Arquitectura de Seguridad Humano-IA

Reporte Transparente (GAMER y SPIRIT-AI)



Obligación de declarar el uso de IA y documentar errores en la investigación médica.

Referencias Bibliográficas de Alto Impacto (Formato Vancouver)

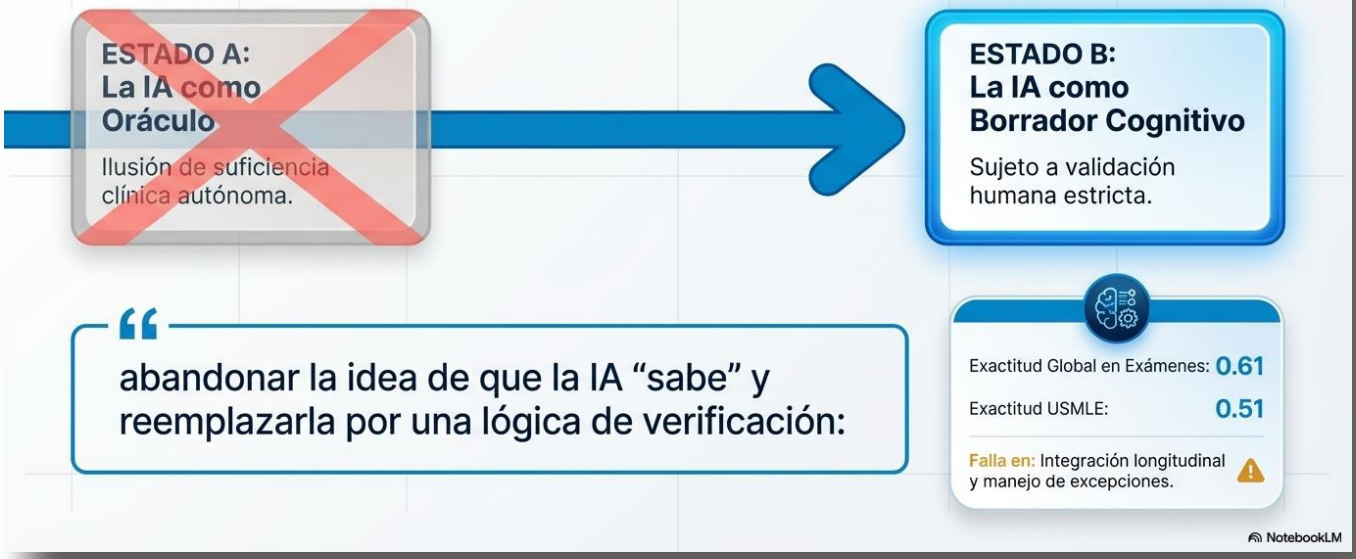
1. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the premising perspectives and valid concerns. Healthcare (Basel). 2022;11(2):887. doi:10.3290/healthcare11680887.
2. HaitanfBerbeide J. Renisch K. The ethice of ChotGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). NPJ Digit Med. 2024;7(1):183. doi:10.1098/e41745-024-01137-4.
3. Omar M. Sorin V. Agbarsia R, Apekema DU, Benjamini M. Shilo S. et al. Evaluatioq and addressing demographic dispatities in medical large language models: a systematic review. Int J Equity Health. 2023-24(1):67. doi:10.1188/s13999-029-03413-0.
4. Luo X, Tham YC, Giuffré M, Renisch K, Caher M, Lam K, et al. Reporting guideline for the use of Generative Artificial intelligence tools in Médical Research: the DAMED Statement. BMJ 29id Raaed Med. 2023;36(6):580-400. doi.10.1130/bngqbm-2023-113223.
5. Abrugongo LM, Maacheroal F, Brouria S, Duarhg S, Senlel J. Damaosi segmentarl gmaoition for large language models in healthcare: a systematic review. PLDS Digit Health. 2023.4(6):e0000677, doi.10.1371/journal pdig 6000677.

Referencia Principal: "Cómo detectar sesgas y alucinaciones en el uso de la inteligencia artificial y cómo exitarlos" (2023).

NotebookLM

La automatización clínica acelera la síntesis pero introduce errores de alta plausibilidad

El modelo de lenguaje genera hipótesis iniciales, no evidencia final



Diagnóstico diferencial de la falla algorítmica: Alucinación vs. Sesgo

ALUCINACIÓN FACTUAL

En términos operativos, "alucinación" debe entenderse como la generación de contenido inexacto, no sustentado o inventado.

Etiología

Falta de información en el modelo, prompts vagos, exigencia de precisión bibliográfica sin repositorios.

Signos Clínicos

- Invención de datos, referencias inexistentes, exageración de conclusiones.

SESGO SISTEMÁTICO

Desviación que perjudica la exactitud, equidad o aplicabilidad en determinados grupos.

Etiología

Replicación de estereotipos del set de entrenamiento, extrapolación de patrones poblacionales a individuos.

Signos Clínicos

- Diferencias de manejo injustificadas por sexo, raza, etnia, idioma o nivel socioeconómico.

NotebookLM

La anatomía de una alucinación bibliográfica indetectable

1. **Autores:** Nombres de investigadores reales extraídos de contextos similares.
2. **Título:** Altamente plausible pero completamente inventado por el modelo.
3. **DOI:** Enlace sintético que arroja error 404 al verificarse en bases de datos.

REGLA CLÍNICA: Ninguna bibliografía generada por un modelo de lenguaje tiene validez sin comprobación directa y manual en bases de datos indexadas (PubMed, Cochrane). Las tasas de citas inventadas son clínicamente inaceptables.

NotebookLM

Matriz de triage para el comportamiento anómalo de la IA

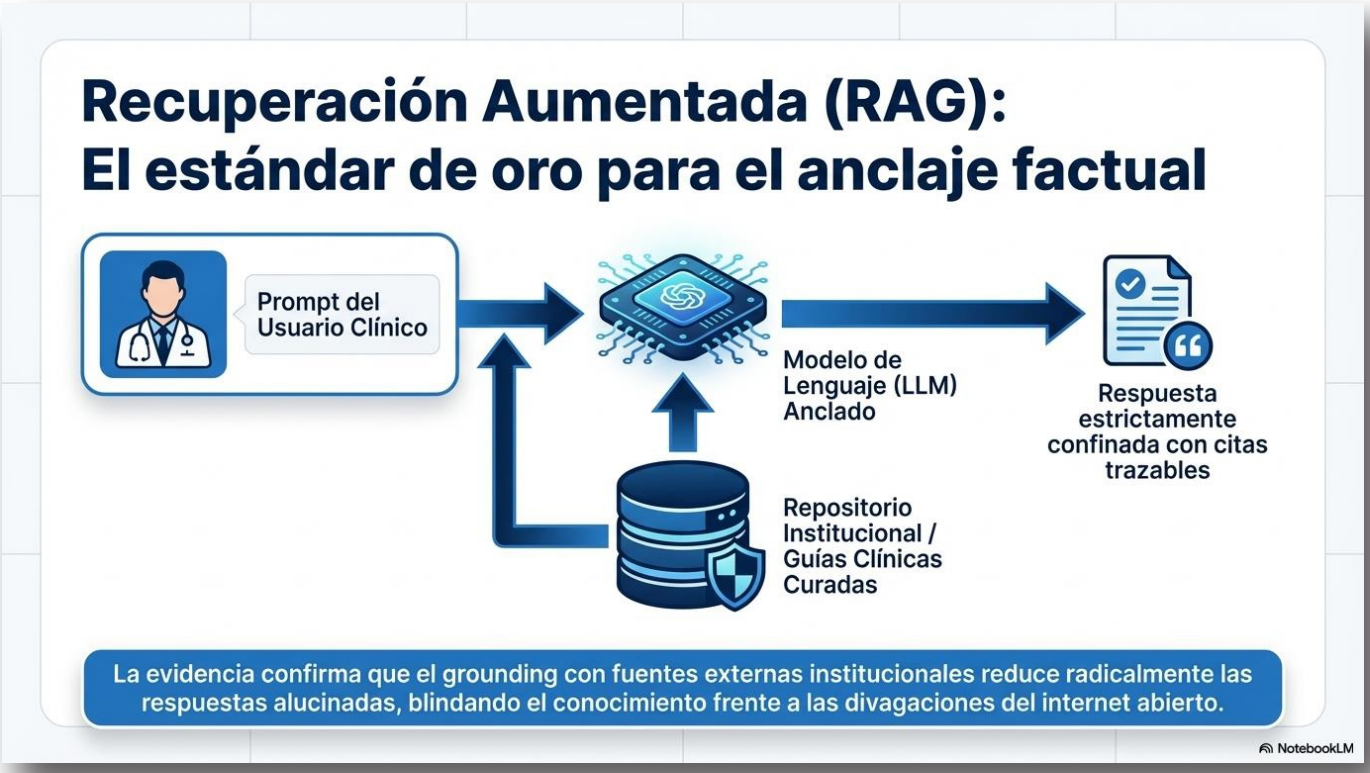
Hallazgo Algorítmico	Patrón Observado	Riesgo Clínico
Alucinación bibliográfica	Referencias con títulos plausibles pero ilocalizables	Sustentar decisiones con evidencia inexistente
Sesgo demográfico	Recomendaciones distintas por sexo/etnia sin justificación	Inequidad diagnóstica o terapéutica
Sobreconfianza algorítmica	Respuesta muy segura ante datos ambiguos o insuficientes	Cierre prematuro del razonamiento médico
Razonamiento frágil	Resumen cosmético que omite limitaciones clave	Error de priorización y falsa certeza

NotebookLM

Cuatro pilares arquitectónicos blindan la autoridad epistémica del clínico

NotebookLM


Checklists de mitigación táctica para la consulta diaria
Paso 1: El Prompt Estructural
Exigir objetivo, contexto, población y formato.
"Responde solo con base en hechos, sin inferencias y dudas"
Beneficio: Reduce deriva semántica.
Paso 2: RAG Institucional
Conectar el modelo exclusivamente a protocolos locales o guías indexadas.
Beneficio: Evita el conocimiento desactualizado de internet.
Paso 3: Doble Lectura Crítica
Usar IA exclusivamente como borrador; la firma y decisión son 100% humanas.
Beneficio: Previene la automatización acrítica.
Paso 4: Auditoría Adversarial
Probar repetidamente el mismo caso clínico cambiando solo la variable de sexo o etnia del paciente.
Beneficio: Destapa sesgos latentes.
NotebookLM


Recuperación Aumentada (RAG): El estándar de oro para el anclaje factual
Prompt del Usuario Clínico
Modelo de Lenguaje (LLM) Anclado
Respuesta estrictamente confinada con citas trazables
Repositorio Institucional / Guías Clínicas Curadas
La evidencia confirma que el grounding con fuentes externas institucionales reduce radicalmente las respuestas alucinadas, blindando el conocimiento frente a las divagaciones del internet abierto.
NotebookLM

Marcos de calidad obligatorios para la investigación médica con IA

Investigación Primaria (Ensayos)

CONSORT-AI y SPIRIT-AI

Obligan a reportar datos de entrada, interacción humano-IA y análisis de fallos.

Evidencia Secundaria

PRISMA 2020 y AMSTAR-2

Esenciales para juzgar la robustez metodológica de revisiones que incluyen IA.

Reporte Transversal

GAMER

Exige declarar con precisión dónde y cómo se usó la IA generativa (extracción, análisis, redacción).

En la medicina basada en evidencia actual, no basta con utilizar la IA; es imperativo reportarla de forma auditable y reproducible.

NotebookLM

El Triage Cognitivo: Repensando el interrogatorio algorítmico

ABANDONAR (Pregunta Pasiva)

~~¿qué respondió la IA?~~

ADOPTAR (El Triage de 3 Vías)

1. ¿de dónde proviene esta afirmación?
2. ¿A qué subgrupo demográfico podría perjudicar esta salida?
3. ¿Qué dato clínico faltante haría cambiar radicalmente esta respuesta?

NotebookLM

Las tres reglas de oro para la implementación clínica de IA

VEREDICTO FINAL: Si la respuesta algorítmica falla en CUALQUIERA de estos tres filtros, debe descartarse inmediatamente o someterse a corrección humana exhaustiva.

NotebookLM

Referencias Bibliográficas de Alto Impacto (I) in Inter Tight Extra Bold, Medical Blue Navy

1. **Sallam M.** ChatGPT utility in healthcare education, research, and practice: systematic review on current state, challenges, and future directions. Healthcare (Basel). 2023;11(6):887. doi:10.3390/healthcare11060887.

2. **Haltaufderheide J, Ranisch R.** The ethics of ChatGPT in medicine and healthcare: a systematic review on large language models (LLMs). NPJ Digit Med. 2024;7(1):183.

3. **Omar M, Sorin V, et al.** Evaluating and addressing demographic disparities in medical large language models to improve equity. Int J Equity Health. 2025;24(1):57.

4. **Waldock WJ, et al.** The accuracy and capability of artificial intelligence solutions in health care examinations: systematic review. J Med Internet Res. 2024;26:e56532.

5. **Fareed M, et al.** A systematic review of ethical considerations of large language models in healthcare: privacy, bias, and transparency. Front Digit Health. 2025.

6. **World Health Organization.** Ethics and governance of artificial intelligence for health. Geneva: WHO; 2021.

NotebookLM

Referencias Bibliográficas de Alto Impacto (II)

7. Chelli M, et al. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews... J Med Internet Res. 2024;26:e53164.

⚠ (Nota de validación: DOI confirmado vía PMID 38776130).

8. Pfohl SR, et al. A toolbox for surfacing health equity harms and biases in large language models. Nat Med. 2024.

9. Page MJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

10. Shea BJ, et al. AMSTAR 2: a critical appraisal tool for systematic reviews... BMJ. 2017;358:j4008.

11. Liu X, et al. Reporting guidelines for clinical trial reports for guidelines or clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020.

Referencias Bibliográficas de Alto Impacto (III)

12. Cruz Rivera S, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020.

13. Luo X, et al. Reporting guideline for the use of Generative Artificial intelligence tools in MEdical Research: the GAMER Statement. BMJ Evid Based Med. 2025.

14. Hake J, et al. Quality, accuracy, and bias in ChatGPT-based summarization of medical abstracts. Oncologist. 2024.

⚠ (Nota de validación: Verificado por PMID 38527823).

15. Amugongo LM, et al. Retrieval augmented generation for large language models in healthcare: a systematic review. PLoS Digit Health. 2025;4(6):e0000877.