



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 3

La inteligencia artificial (IA) en el ámbito médico no es un fenómeno emergente derivado de la reciente popularidad de modelos generativos como ChatGPT. Su trayectoria es una evolución continua que hunde sus raíces en la lógica formal y la computación del siglo XX. Sin embargo, nos encontramos en un punto de inflexión donde la transición no es únicamente técnica, sino profundamente lingüística y epistemológica. Para la comunidad médica académica, la adopción de estas herramientas exige una alfabetización crítica que permita distinguir entre la innovación técnica y la evidencia clínica real.

Evolución Histórica de la IA Médica

El desarrollo de la IA en medicina se puede estructurar en cuatro periodos fundamentales, cada uno con un núcleo tecnológico y un aporte clínico distintivo. La comprensión de este proceso permite evitar la "ilusión de novedad absoluta" y situar las herramientas contemporáneas en su contexto adecuado.

Tabla 1. Periodos de la IA en Medicina y su impacto terminológico

Periodo	Núcleo Tecnológico	Aporte Médico Principal	Lenguaje Característico
Era Simbólica	Reglas, lógica, sistemas expertos.	Ayuda diagnóstica basada en conocimiento explícito.	Regla, inferencia, sistema experto, base de conocimiento.
Era Estadística	Modelos probabilísticos, regresión, minería de datos.	Estratificación de riesgo y apoyo cuantitativo.	Probabilidad, riesgo, predictor, validación.
Era de Aprendizaje Profundo	Redes neuronales, visión computacional, NLP.	Imagen médica, clasificación, reconocimiento de patrones.	Red neuronal, deep learning, entrenamiento, sobreajuste.
Era Generativa / Fundacional	LLM, transformers, multimodalidad.	Resumen clínico, documentación, educación, interacción textual.	Prompt, contexto, alucinación, grounding, explicabilidad, trazabilidad.

El Estado Actual de la Evidencia Clínica

A pesar de la acelerada velocidad de publicación, la madurez clínica de la IA sigue siendo desigual. Las revisiones sistemáticas recientes subrayan una brecha crítica: la mayoría de las evaluaciones se centran en tareas técnicas o artificiales (como preguntas de exámenes médicos) y no en desenlaces clínicos reales.

- Evidencia en Modelos de Lenguaje (LLM):** Solo el 5% de los estudios sobre aplicaciones sanitarias de LLM utilizan datos reales de atención. Persisten debilidades en la evaluación de sesgo, toxicidad, calibración e incertidumbre.
- Dominios de Aplicación:** Se han identificado tres áreas principales de intervención: apoyo a la decisión clínica, optimización organizacional y comunicación clínica.
- Documentación y Educación:** Existe un crecimiento notable en el resumen automatizado de textos (especialmente en UCI y Radiología) y en la generación de material educativo para pacientes. No obstante, se advierten problemas de legibilidad, exactitud y variabilidad en las respuestas.

El Nuevo Lenguaje Médico: Una Necesidad Operativa

La IA ha introducido nuevos objetos de trabajo que obligan a revisar conceptos tradicionales como diagnóstico, pronóstico y autonomía. El uso impreciso de términos como "predicción" o "validación" puede inducir a errores conceptuales y metodológicos graves.

Tabla 2. Glosario Crítico para el Médico Contemporáneo

Término	Definición Clínica Útil	Riesgo de Mal Uso / Confusión
IA Generativa	Modelos que crean texto, imagen u otros contenidos originales.	Equiparar la fluidez del lenguaje con la veracidad clínica.
Validación Externa	Prueba del modelo en una población distinta a la de entrenamiento.	Llamar "validado" a un modelo probado solo internamente.
Calibración	Concordancia entre la probabilidad predicha y la frecuencia observada.	Reducir el rendimiento únicamente a la exactitud o al área bajo la curva (AUC).
Alucinación	Generación textual no sustentada o fuera de contexto.	Interpretarla solo como una equivocación menor; es un error estructural.
Explicabilidad	Capacidad de ofrecer razones comprensibles sobre una salida o decisión.	Confundir una visualización gráfica con una explicación suficiente.
Robustez	Estabilidad del desempeño ante cambios razonables de contexto.	Ignorar la fragilidad del modelo fuera de su entorno original.
Sesgo	Distorsión sistemática que afecta la validez o la equidad.	Usarlo como rótulo genérico para cualquier error aleatorio.
Trazabilidad	Posibilidad de rastrear datos, versiones de modelos y decisiones.	Omitir la versión del modelo o la fuente de los datos de entrenamiento.

Implicaciones para la Práctica y la Academia

Para una institución como la Academia Nacional de Medicina de México, el desafío reside en la "alfabetización crítica". Esto implica migrar de una lectura centrada en "qué tan convincente suena" la IA, hacia una centrada en "cómo fue validada".

Directrices Metodológicas y Éticas

Es imperativo que la comunidad médica domine y exija el cumplimiento de marcos internacionales de reporte y gobernanza:

1. **SPIRIT-AI y CONSORT-AI:** Para ensayos clínicos que involucran IA.
2. **PRISMA 2020 y AMSTAR 2:** Para la valoración de revisiones sistemáticas.
3. **GAMER Statement:** Para reportar el uso de IA generativa en la investigación médica (exigiendo transparencia en prompts y verificación humana).
4. **Guía Ética de la OMS:** Para asegurar que la gobernanza de la IA priorice la equidad, la seguridad y la privacidad.

El Juicio Clínico en la Era Digital

La IA no reemplaza el razonamiento médico, pero redefine sus condiciones. El médico ya no solo "usa software", sino que "dialoga con modelos". Por ello, el clínico del presente debe aprender a realizar preguntas de control críticas:

- ¿Con qué datos externos se validó este modelo?
- ¿Qué tan calibrado está para mi población específica?
- ¿Qué sesgos muestra y quién supervisa su uso?
- ¿Qué parte del proceso sigue siendo indelegable para el médico?

Conclusiones

La IA médica ha entrado en una fase lingüísticamente decisiva. Hablar correctamente de IA es hoy tan fundamental como usarla correctamente. Para la medicina académica, la prioridad debe ser garantizar que la innovación no se confunda con evidencia y que la automatización no desplace al juicio clínico riguroso. La IA debe diseñarse y desplegarse bajo principios de trazabilidad, robustez y transparencia para asegurar desenlaces que sean verdaderamente relevantes para el paciente.

Referencias Bibliográficas

1. 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.
2. 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 randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008. doi:10.1136/bmj.j4008.
3. Ibrahim H, Liu X, Cruz Rivera S, Moher D, Chan AW, Sydes MR, et al. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. *Trials*. 2021;22(1):11. doi:10.1186/s13063-020-04951-6.
4. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.
5. Wilhelm C, Steckelberg A, Rebitschek FG. Benefits and harms associated with the use of AI-related algorithmic decision-making systems by healthcare professionals: a systematic review. *Lancet Reg Health Eur*. 2024;48:101145. doi:10.1016/j.lanepe.2024.101145.
6. Melnyk O, et al. Generative Artificial Intelligence Terminology: A Primer for Clinicians and Medical Researchers. *Cureus*. 2023. doi:10.7759/cureus.53187.

7. Luo X, Tham YC, Giuffrè M, Ranisch R, Daher M, Lam K, et al; GAMER Working Group. 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.
8. Grossi E. The long journey of artificial intelligence in medicine: an overview. *Clin Exp Rheumatol*. 2025;43(5):815-821. doi:10.55563/clinexprheumatol/oamfed.
9. Bedi S, Callahan A, Shah NH, et al. Testing and Evaluation of Health Care Applications of Large Language Models: A Systematic Review. *JAMA Netw Open*. 2025;10. Aydin S, Karabacak M, Vlachos K, Margetis K. Large language models in patient education: a scoping review of applications in medicine. *Front Med (Lausanne)*. 2024;11:1477898. doi:10.3389/fmed.2024.1477898.
11. Bednarczyk L, Reichenpfader D, Gaudet-Blavignac C, Ette AK, Zaghir J, Zheng Y, et al. Scientific Evidence for Clinical Text Summarization Using Large Language Models: Scoping Review. *J Med Internet Res*. 2025;27:e68998. doi:10.2196/68998.
12. Lekadir K, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ*. 2025;388:r340. doi:10.1136/bmj.r340.
13. Khosravi M, Zare Z, Mojtabaeian SM, Izadi R. Artificial Intelligence and Decision-Making in Healthcare: A Thematic Analysis of a Systematic Review of Reviews. *Health Serv Res Manag Epidemiol*. 2024;11:23333928241234863. doi:10.1177/23333928241234863.



Historia y Nuevo Lenguaje de la IA en Medicina

Una revisión analítica y marco epistemológico para clínicos

Presentado ante la Academia Nacional de Medicina de México



NotebookLM

De la Lógica a la Generación: La Nueva Semiotología de la IA Médica

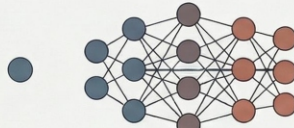
Evolución Histórica de la IA en Medicina

La IA médica es una continuidad histórica que ha transitado desde sistemas basados en reglas hasta modelos generativos. Esta evolución exige que el médico moderno adopte un nuevo lenguaje técnico y ético para no confundir la fluidez textual con la veracidad clínica ni la innovación con evidencia.



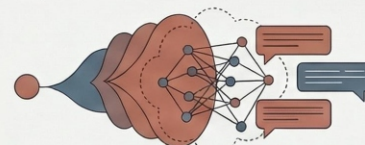
Era Simbólica y Estadística (Pre-2010s)

Transición de sistemas expertos basados en reglas a modelos probabilísticos para estratificación de riesgo.



Auge del Aprendizaje Profundo (2012-2022)

Dominio de redes neuronales aplicadas a visión computacional, imágenes médicas y reconocimiento de patrones.



Era Generativa y Fundacional (2022-Presente)

Irrupción de LLMs y modelos multimodales que democratizan el acceso mediante interfaces conversacionales.

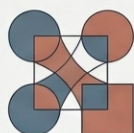
El Nuevo Lenguaje del Juicio Clínico

Validación Externa y Calibración



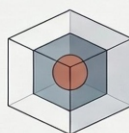
Evaluación del modelo en poblaciones distintas al entrenamiento para asegurar concordancia entre predicción y realidad.

Alucinación vs. Veracidad Clínica



Reconocer que la fluidez lingüística de un LLM no garantiza una inferencia médica válida.

Explicabilidad, Sesgo y Trazabilidad



Capacidad de rastrear decisiones y detectar distorsiones sistemáticas que afecten la equidad del paciente.

Resumen de Evaluación Clínica

Dimensión Evaluada	Estado en Literatura Reciente
Datos de Atención Real	Solo el 5% de los estudios los utilizan
Enfoque de Pruebas	Predominan preguntas de examen sobre escenarios reales
Medición de Desenlaces	Evidencia limitada en beneficios directos para el paciente

1. Page MJ, Moher D, Altman DG, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
2. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guideline. Geneva: WHO; 2021. 3. Williams G, Szallasi Z, et al. Benefits and harms associated with use of artificial intelligence decision-making systems by healthcare professionals: a systematic review. *Lancet Reg Health Eur*. 2022;68-70:1010. 4. Kohn M, et al. PUTTING IT ALL TOGETHER: International consensus guideline for trustworthy and deployable artificial intelligence healthcare. *BMJ*. 2022;376:n400. 5. Kim H, Doherty D, et al. Reporting guideline for the use of Generative Artificial Intelligence tools in Medical Research: the GAINP System. *BMJ End Metab*. 2023;3(3):199-209.

NotebookLM



La Promesa



La Realidad

El Paradigma Actual en Medicina Digital

- La IA en medicina no nació con ChatGPT ni con los modelos generativos.
- Existe una **brecha crítica**: la implementación clínica sigue retrasada frente a la velocidad de publicación.
- El desafío central ya no es de código, sino epistemológico: evaluar rigurosamente herramientas con madurez clínica desigual.
- Se requiere con **urgencia** un nuevo marco clínico y de lenguaje para no confundir innovación con evidencia.

NotebookLM

La IA no nació hoy: Las Cuatro Eras



Era Simbólica: Reglas, Reglas, lógica y sistemas expertos.
Lenguaje: Inferencia, base de conocimiento.



Era Estadística: Modelos probabilísticos y minería de datos.
Lenguaje: Probabilidad, riesgo, predictor.



Aprendizaje Profundo: Redes neuronales aplicadas a imagen médica y señales.
Lenguaje: Entrenamiento, sobreajuste.



Era Generativa: Modelos fundacionales y LLMs.
Lenguaje: Prompt, contexto, alucinación, trazabilidad.

La verdadera novedad actual no es la IA, sino la masificación de las **interfaces lingüísticas**. El médico pasó de usar software a **dialogar con modelos**.

NotebookLM

El Espejismo de la Interfaz Generativa

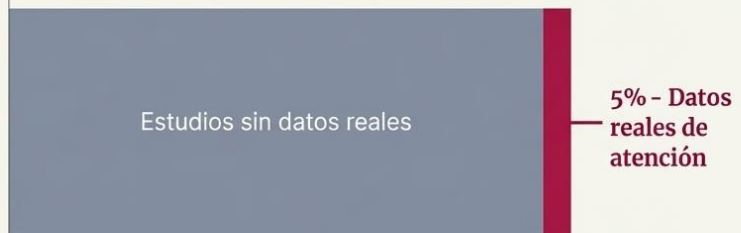
- **El Riesgo Epistemológico:** La persuasión aumenta la utilidad, pero también la posibilidad de **sobreconfianza**.
- **La Trampa de la Fluidez:** Un texto bien redactado por un Gran Modelo de Lenguaje (LLM) **no equivale a una inferencia clínica válida**.
- **El Cambio Necesario:** Debemos migrar de una lectura centrada en qué tan convincente suena a otra **centrada en cómo fue validado y en qué población**.



NotebookLM

Aplicaciones Clínicas frente a Evidencia Real

- **Dominios de Aplicación:** Apoyo a la decisión clínica, optimización organizacional y comunicación/documentación.
- **El Dato Crítico:** Una revisión sistemática reciente sobre LLMs en salud encontró que solo el **5%** de los estudios utilizaron datos reales de atención.
- **La Realidad de la Literatura:** La mayoría de la evidencia actual evalúa el desempeño del modelo en exámenes médicos (tareas artificiales), subevaluando el impacto real en desenlaces centrados en el paciente, sesgos y toxicidad.



NotebookLM

El Problema Semántico en Medicina Digital

~~Predicción~~



Error Común 1: Llamar predictivo a un análisis retrospectivo.

Un error menor pero significativo: llamar "predictivo" a un modelo que en realidad solo analiza datos históricos.

~~Asociación~~



Error Común 2: Confundir asociación estadística con **predicción** de eventos en datos no vistos.

~~IA~~



Error Común 3: Usar IA como sinónimo genérico para cualquier automatización de procesos.

Advertencia: Este desorden conceptual contamina la investigación y la evaluación de tecnologías.

- El desorden conceptual contamina la investigación clínica y la adopción institucional.

NotebookLM

El Nuevo Lenguaje Médico: El Núcleo Tecnológico

Aprendizaje Automático (ML)

Definición: Aprende patrones a partir de datos.

Riesgo de mal uso: Confundirlo con causalidad clínica.

Aprendizaje Profundo (Deep Learning)

Definición: ML con redes neuronales multicapa.

Riesgo de mal uso: Asumir superioridad universal sobre modelos más simples.

Gran Modelo de Lenguaje (LLM)

Definición: Modelo entrenado en corpus masivos para generar lenguaje.

Riesgo de mal uso: Interpretarlo como un **médico virtual confiable** por defecto.

IA Generativa

Definición: Crea texto, imagen u otros contenidos nuevos.

Riesgo de mal uso: Equiparar fluidez de texto con **veracidad médica**.

Prompt

Definición: Instrucción o contexto dado al modelo.

Riesgo de mal uso: Creer que un buen prompt **corrige la falta de validación del modelo**.

NotebookLM

El Nuevo Lenguaje Médico: Seguridad y Evaluación

Validación Externa

Definición: Prueba en una población distinta a la de entrenamiento.

Riesgo de mal uso: Llamar validado a un modelo solo probado internamente.

Calibración

Definición: Concordancia entre probabilidad predicha y frecuencia observada.

Riesgo de mal uso: Reducir el rendimiento solo al AUC o exactitud.

Explicabilidad

Definición: Capacidad de ofrecer razones comprensibles.

Riesgo de mal uso: Confundir cualquier visualización con explicación suficiente.

Sesgo

Definición: Distorsión sistemática que afecta validez o equidad.

Riesgo de mal uso: Usarlo como rótulo genérico para cualquier error aleatorio.

Robustez

Definición: Estabilidad del desempeño ante variaciones razonables.

Riesgo de mal uso: Ignorar la fragilidad del modelo fuera de su entorno original.

NotebookLM

La Evolución de la Pregunta Clínica

La pregunta tradicional:
¿Sirve este software?

La nueva semiología clínica exige preguntar:

- ¿Qué modelo es y cómo fue entrenado?
- ¿Con qué datos externos **se validó**?
- ¿Qué tan **calibrado** está para nuestra población?
- ¿En qué falla y qué **sesgos** muestra?
- ¿Quién **supervisa** su uso y qué parte del proceso **sigue siendo indelegable para el médico**?

NotebookLM

Mensajes Clave para la Academia Nacional de Medicina

1. **Histórico:** La IA representa una **continuidad clínica**, no una ruptura absoluta.

2. **Epistemológico:** En IA generativa, la fluidez del lenguaje no es sinónimo de verdad clínica.

3. **Lingüístico:** Exigir un vocabulario operativo y riguroso; la **claridad terminológica** es innegociable.

4. **Ético:** La trazabilidad, privacidad y responsabilidad exigen marcos de **gobernanza institucional**.

5. **Práctico:** La investigación y publicación académica requieren una estricta adopción de **estándares metodológicos internacionales**.

NotebookLM

Gobernanza y Marcos de Reporte en Publicaciones

Para evitar la confusión entre innovación y evidencia, se exige adherencia a:



**PRISMA 2020 /
AMSTAR 2**

Obligatorios para revisiones sistemáticas de IA.



**SPIRIT-AI /
CONSORT-AI**

Estándares para el reporte de ensayos clínicos que incluyen intervenciones de IA.



**GAMER Statement
(2025)**

Exige 9 dominios mínimos para reportar el uso de IA generativa en investigación médica.



**WHO Guidance /
FUTURE-AI**

Directrices para el diseño ético, confiable y equitativo de IA en la salud pública.

NotebookLM



La Redefinición del Juicio Clínico

“La IA no reemplaza el juicio clínico, pero sí **redefine las condiciones** bajo las cuales ese juicio debe **ejercerse**. En esta nueva etapa académica, **hablar correctamente de IA** será casi tan **importante** como **usarla correctamente**.”

NotebookLM

Referencias Bibliográficas

1. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.
2. Luo X, Tham YC, Giuffrè M, Ranisch R, Daher M, Lam K, et al; GAMER Working Group. 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.
3. 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.
4. Ibrahim H, Liu X, Cruz Rivera S, Moher D, Chan AW, Sydes MR, et al. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. Trials. 2021;22(1):11.
5. Lekadir K, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:r340.

NotebookLM