

Inteligencia artificial generativa y evidencia educativa: el desafío de una tecnología que cambia antes de publicarse

Generative artificial intelligence and educational evidence: the challenge of a technology that changes before publication

“La exigencia de objetividad científica hace inevitable que toda afirmación científica sea tentativa para siempre”.

KARL POPPER

La inteligencia artificial generativa (IAGen) ha pasado, en un lapso extraordinariamente breve, de ser una novedad tecnológica a convertirse en objeto formal de indización biomédica. La *National Library of Medicine* ya reconoce “Generative Artificial Intelligence” como descriptor MeSH (*Medical Subject Headings*) a partir del año 2025 (<https://meshb.nlm.nih.gov/record/ui?ui=D000098842>). El dato parece solo una curiosidad técnica o bibliotecológica, pero no lo es. Cuando un fenómeno alcanza la categoría de término MeSH, deja de ser solo conversación pública o moda industrial: entra al vocabulario controlado con el que MEDLINE ordena la evidencia. Para quienes trabajamos en educación en profesiones de la salud, el mensaje es claro; la IAGen ya forma parte del lenguaje estable de la medicina, aunque el objeto que nombra esté lejos de haberse estabilizado¹.

La paradoja es aparente: la nomenclatura comienza a fijarse justo cuando la tecnología se acelera. Carl Hendrick ha descrito con crudeza que una parte importante de la investigación educativa sobre IA nace “muerta al llegar”, no por falta de rigor, sino porque estudia versiones de sistemas que ya fueron superadas cuando el artículo aparece². En su argumento, un estudio realizado con GPT-3.5 en 2023 puede atravesar recolección, revisión y arbitraje durante muchos meses y publicarse cuando el ecosistema ya cambió varias veces. La preocupación no es solo retórica. METR reportó que la longitud de las tareas que los modelos frontera pueden completar de manera autónoma se ha venido duplicando aproximadamente cada siete meses (<https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/>). En un entorno así, la validez temporal se vuelve tan importante como la validez interna.

En educación en profesiones de la salud no partimos de cero en este problema. Un estudio sobre revistas de *health professions education* documentó un promedio de 180.9 días desde el envío hasta la aceptación y de 263.6 días desde el envío hasta la aparición en PubMed³. Incluso sin hablar de dos

años completos, nueve meses bastan para que un manuscrito sobre IAGen envejezca de manera sustantiva. En otras palabras, no siempre estamos frente a evidencia falsa; muchas veces estamos frente a evidencia rezagada. Y una evidencia rezagada puede conducir a decisiones curriculares, regulatorias o de política editorial mal calibradas para el presente y el futuro cercano.

Por eso el hecho de que IAGen sea ya un término MeSH tiene una doble lectura. Es una buena noticia porque mejora la recuperabilidad y la trazabilidad de la literatura. Pero también es una advertencia metodológica: las búsquedas exhaustivas no pueden apoyarse solo en el descriptor nuevo. Si buena parte de los trabajos de 2023 y 2024 quedaron bajo un término paraguas más amplio, las revisiones serias tendrán que combinar MeSH con texto libre, sinónimos y nombres de herramientas. Quien busque únicamente “Generative Artificial Intelligence” corre el riesgo de construir una falsa sensación de exhaustividad y pasar por alto la literatura temprana que dio forma al campo.

El segundo fenómeno que reconfigura el paisaje es el de los *preprints*. En educación médica ya no son una rareza. Maggio y colegas identificaron 204 *preprints* clasificados como “*medical education*” en medRxiv entre 2019 y 2022; en promedio fueron descargados 1,875 veces, todos circularon en redes sociales y casi la mitad terminó publicándose después en revistas arbitradas⁴. Es decir, el *preprint* no es solo un borrador invisible: es un documento que circula, se comenta, orienta conversaciones académicas y, en no pocos casos, anticipa la versión final de trabajos que después se formalizan en el circuito editorial tradicional. Ignorarlo no elimina su influencia; solo nos deja menos preparados para leerla críticamente.

Pero reconocer su influencia no equivale a confundir estatutos epistemológicos. medRxiv lo dice claramente: los *preprints* son reportes preliminares que no han sido certificados por revisión por pares y no deben usarse para guiar la práctica clínica. Ese principio debe extrapolarse con prudencia a la educación médica: un *preprint* puede alimentar deliberación, pero no debería traducirse por sí mismo en rediseño curricular, políticas de evaluación o

recomendaciones institucionales. Al mismo tiempo, tampoco conviene caer en el extremo opuesto. Nelson y colaboradores, al comparar 100 *preprints* con sus versiones publicadas, encontraron cambios promedio de 6% en las estimaciones puntuales, una correlación muy alta entre versiones y una reducción de 7% en la amplitud de los intervalos de confianza tras la revisión⁵. El mensaje es matizado: los *preprints* no son equivalentes a artículos finales, pero tampoco son simple ruido en la conversación de la investigación científica.

De hecho, la propia *National Library of Medicine* ha reconocido que la velocidad importa. El *NIH Preprint Pilot* busca aumentar la visibilidad de resultados tempranos financiados por NIH a través de PubMed Central y, por extensión, PubMed (<https://pmc.ncbi.nlm.nih.gov/about/nihpreprints/>). Conviene subrayar la distinción: que un *preprint* sea recuperable en PubMed no lo convierte en registro MEDLINE ni en evidencia ya arbitrada. Sin embargo, sí transforma la ecología de búsqueda. La frontera entre literatura “formal” y literatura “en circulación” se vuelve más porosa, y el lector necesita competencias nuevas para no mezclar descubribilidad con validación.

En este contexto, mantenerse actualizado exige una estrategia de lectura por capas. La literatura arbitrada sigue siendo indispensable para preguntas de mayor vida media: marcos conceptuales, ética, gobernanza, efectos pedagógicos relativamente estables y métodos de evaluación. Para preguntas de vida media corta, por ejemplo, qué puede hacer hoy un modelo específico, con qué interfaz, bajo qué condiciones de acceso y con qué nivel de supervisión, los *preprints*, reportes técnicos y análisis rápidos pueden aportar información más vigente que el artículo tradicional. Pero esa vigencia solo sirve si va acompañada de lectura crítica. Dijkstra y colegas han recordado que leer un artículo con IA implica preguntar por la tarea analizada, los datos, el contexto de uso, la participación humana, los sesgos, la explicabilidad y la transparencia del reporte⁶.

La educación en profesiones de la salud ofrece una lección adicional, el campo comenzó a producir síntesis secundarias antes de alcanzar estabilidad tecnológica. Un metaanálisis de ensayos aleatoriza-

dos sobre enseñanza basada en IAGen en estudiantes de medicina halló que, en conjunto, no había diferencias significativas en adquisición teórica de conocimientos frente a métodos tradicionales, aunque sí mejores resultados en desarrollo de habilidades prácticas y mayor satisfacción estudiantil⁷. Este tipo de hallazgos es valioso, pero exige una lectura prudente. Cuando decimos “IAGen” no describimos una intervención unitaria; describimos una familia cambiante de modelos, interfaces, ventanas contextuales, políticas de acceso y modalidades de uso. Reificar la categoría como si fuera estable puede llevar a conclusiones más sólidas en apariencia que en realidad.

Desde la perspectiva editorial, esto obliga a elevar los estándares de descripción metodológica. No basta con escribir “se utilizó ChatGPT”. Deberíamos pedir, de forma rutinaria, el nombre del modelo, la versión, la fecha exacta de uso o de recolección de datos, la modalidad de acceso, el idioma, la estrategia de *prompting*, la presencia o ausencia de verificación humana y cualquier actualización relevante ocurrida durante el estudio. Hendrick ha propuesto algo semejante a un “*technology timestamp*”, y la idea es interesante: en investigación sobre IAGen, el tiempo no es un dato administrativo; es una variable científica². Sin esa marca temporal, el lector no sabe si está frente a un hallazgo vigente o a una reconstrucción arqueológica de una herramienta ya desplazada.

También necesitamos formatos editoriales más ágiles. Las revisiones vivas, las comunicaciones breves, las actualizaciones metodológicas y los observatorios de evidencia pueden resultar más útiles que la ilusión de una síntesis definitiva. Tong y colaboradores han propuesto métodos para incorporar *preprints* a revisiones sistemáticas sin renunciar por completo al control de sesgos; ese esfuerzo apunta en la dirección correcta⁸. No se trata de sacrificar rigor por velocidad, sino de reconocer que, en determinados dominios, la demora también produce error. La lentitud no siempre es sinónimo de prudencia; a veces es simplemente desactualización institucionalizada.

Que IAGen haya ingresado al tesoro MeSH es, simultáneamente, un hito y una llamada de atención. El hito es que la inteligencia artificial generativa ya

pertenece al repertorio nominal de la ciencia biomédica. La llamada de atención es que la aparición de un nombre estable no garantiza un objeto estable. Para quienes investigamos, enseñamos y editamos en educación en profesiones de la salud, el desafío no consiste en elegir entre el artículo arbitrado y el *preprint*, entre la cautela y la rapidez, o entre la fascinación y el rechazo. El desafío real es construir una cultura de juicio crítico capaz de distinguir entre evidencia sólida, evidencia provisional y evidencia obsoleta. En la era de la IAGen, mantenerse actualizado ya no significa leer más: significa leer mejor, leer en tiempo y leer con memoria metodológica. 🔍



Melchor Sánchez Mendiola

EDITOR

Facultad de Medicina, UNAM

REFERENCIAS

1. Zhao Y, Kingston N, Ginsberg R. The death and rebirth of research in education in the age of AI: problems and promises. *ECNU Rev Educ*. 2026;9(1):1-24. Disponible en: <https://doi.org/10.1177/20965311251367053>
2. Hendrick C. Dead on Arrival: AI Is Changing Faster Than Science Can Study It [Internet]. LinkedIn; 2026 Feb 14 [citado Mar 8 2026]. Disponible en: <https://www.linkedin.com/pulse/dead-arrival-ai-changing-faster-than-science-can-study-carl-hendrick-8rbwe>
3. Maggio LA, Bynum WE, Schreiber-Gregory DN, Durning SJ, Artino AR Jr. When will I get my paper back? A replication study of publication timelines for health professions education research. *Perspect Med Educ*. 2020;9(3):139-146. Disponible en: <https://doi.org/10.1007/s40037-020-00576-2>
4. Maggio LA, Costello JA, Artino AR Jr. Describing the landscape of medical education preprints on MedRxiv: current trends and future recommendations. *Acad Med*. 2024;99(9):981-986. Disponible en: <https://doi.org/10.1097/ACM.00000000000005742>
5. Nelson L, Ye H, Schwenn A, Lee S, Arabi S, Hutchins BI. Robustness of evidence reported in preprints during peer review. *Lancet Glob Health*. 2022;10(11):e1684-e1687. Disponible en: [https://doi.org/10.1016/S2214-109X\(22\)00368-0](https://doi.org/10.1016/S2214-109X(22)00368-0)
6. Dijkstra P, Greenhalgh T, Mekki YM, Morley J. How to read a paper involving artificial intelligence (AI). *BMJ Med*.

2025;4:e001394. Disponible en: <https://doi.org/10.1136/bmjmed-2025-001394>

7. Li J, Yin K, Wang Y, Jiang X, Chen D. Effectiveness of generative artificial intelligence-based teaching versus traditional teaching methods in medical education: a meta-analysis of randomized controlled trials. *BMC Med Educ.* 2025;25(1):1175. Disponible en: <https://doi.org/10.1186/s12909-025-07750-2>
8. Tong J, Sun Y, Hubbard RA, Saine ME, Xu H, Zuo X, et al. Incorporating preprints in systematic reviews: a preliminary study of a novel method for rapid evidence synthesis. *J Am Med Inform Assoc.* 2025;32(11):1654-1663. Disponible en: <https://doi.org/10.1093/jamia/ocaf111>