



INTELIGENCIA ARTIFICIAL GENERATIVA Y MODELOS DE LENGUAJE DE GRAN TAMAÑO COMO HERRAMIENTAS DE APOYO METODOLÓGICO EN LA INVESTIGACIÓN EN SALUD: UN ANÁLISIS BIBLIOMÉTRICO

INTELIGÊNCIA ARTIFICIAL GENERATIVA E MODELOS DE LINGUAGEM DE GRANDE ESCALA COMO FERRAMENTAS DE APOIO METODOLÓGICO NA PESQUISA EM SAÚDE: UMA ANÁLISE BIBLIOMÉTRICA

GENERATIVE ARTIFICIAL INTELLIGENCE AND LARGE LANGUAGE MODELS AS METHODOLOGICAL SUPPORT TOOLS IN HEALTH RESEARCH: A BIBLIOMETRIC ANALYSIS

Yasmin Saba de Almeida¹
Elaine Antunes Cortez¹
Geilsa Soraia Cavalcanti Valente¹

ORCID: 0000-0003-2391-7009
ORCID: 0000-0003-3912-9648
ORCID: 0000-0003-4488-4912

¹ Universidade Federal Fluminense, Escola de Enfermagem Aurora de Afonso Costa, Niterói, Rio de Janeiro, Brasil.

Cómo citar: Almeida YS, Cortez EA, Valente GSC. Generative Artificial Intelligence and Large Language Models as methodological support tools in health research: a bibliometric analysis. Online Braz J Nurs. 2026;25(Suppl1):e20267204. <https://doi.org/10.17665/1676-4285.20267204>

RESUMO

Objetivo: Mapear a produção científica e as tendências temáticas do uso de inteligência artificial generativa (GenAI) e modelos de linguagem de grande escala (LLMs) como ferramentas de apoio metodológico na pesquisa em saúde. **Método:** Estudo bibliométrico realizado em setembro de 2026 nas bases PubMed, Scopus e Web of Science. Os registros foram triados por meio de um pipeline híbrido (humano-IA) com apoio de um LLM segundo critérios previamente definidos e revisão humana. Os dados foram analisados no RStudio com uso do pacote *bibliometrix* e de sua interface *biblioshiny*. **Resultados:** O corpus final compreendeu 552 documentos, distribuídos em 337 fontes, com idade média de 0,88 anos e taxa anual de crescimento de 70,7 %. Os Estados Unidos concentraram o maior número de afiliações de autores, seguidos pela China. A literatura apresentou alta dispersão entre periódicos e forte concentração em autores ocasionais (89,2 % com apenas uma publicação), além de modesta colaboração internacional (14,7 %). O mapa temático revelou que os temas motores permanecem centrados no uso genérico de LLMs e ChatGPT, enquanto aplicações metodológicas específicas ainda não constituem agrupamentos consolidados. **Conclusão:** O uso da GenAI e dos LLMs como apoio metodológico à pesquisa em saúde constitui um campo emergente e em rápida expansão, mas ainda desprovido de uma estrutura temática e de canais de publicação consolidados. A consolidação da área exige o estabelecimento de critérios normativos de validação, transparência no relato, supervisão humana estrita e letramento dos pesquisadores em inteligência artificial, garantindo a integridade das evidências científicas produzidas, com repercussões diretas na qualificação das práticas e na segurança do cuidado em saúde.

Descritores: Bibliometria; Inteligência Artificial Generativa; Modelos de Linguagem de Grande Escala; Saúde; Técnicas de Pesquisa.

ABSTRACT

Objective: To map the scientific output and thematic trends in the use of generative artificial intelligence (GenAI) and large language models (LLMs) as methodological support tools in health research. **Method:** Bibliometric study conducted in September 2026 in the PubMed, Scopus, and Web of Science databases. Records were screened through a hybrid (human-AI) pipeline, in which an LLM applied predefined criteria and a human reviewer checked the results. Data were analyzed in RStudio using the *bibliometrix* package and its *biblioshiny* interface. **Results:** The final corpus comprised 552 documents from 337 sources, with a mean document age of 0.88 years and an annual growth rate of 70.7 %. The United States accounted for the largest number of author affiliations, followed by China. The literature showed high dispersion across journals and a strong concentration of occasional authors (89.2 % with a single publication), along with modest international collaboration (14.7 %). The thematic map revealed that motor themes remain centered on the generic use of LLMs and ChatGPT, whereas specific methodological applications have not yet formed consolidated clusters. **Conclusion:** The use of GenAI and LLMs as methodological support in health research is an emerging and rapidly expanding field, but it still lacks a consolidated thematic structure and established publication venues. Consolidating the field will require normative validation criteria, transparent reporting, strict human oversight, and AI literacy among researchers, so as to safeguard the integrity of the scientific evidence produced, with direct implications for the quality of practice and the safety of health care.

Descriptors: Bibliometrics; Generative Artificial Intelligence; Large Language Models; Health; Investigative Techniques.

RESUMEN

Objetivo: Mapear la producción científica y las tendencias temáticas del uso de la inteligencia artificial generativa (GenAI) y de los modelos de lenguaje de gran tamaño (LLM) como herramientas de apoyo metodológico en la investigación en salud. **Método:** Estudio bibliométrico realizado en septiembre de 2026 en las bases de datos PubMed, Scopus y Web of Science. Los registros se cribaron mediante un flujo híbrido (humano-IA) en el que un LLM aplicó criterios definidos previamente y una revisora humana verificó los resultados. Los datos se analizaron en RStudio con el paquete *bibliometrix* y su interfaz *biblioshiny*. **Resultados:** El corpus final estuvo compuesto por 552 documentos, distribuidos en 337 fuentes, con una antigüedad media de 0,88 años y una tasa de crecimiento anual del 70,7 %. Estados Unidos concentró el mayor número de afiliaciones de autores, seguido de China. La literatura mostró una alta dispersión entre revistas y una fuerte concentración de autores ocasionales (el 89,2 % con una sola publicación), además de una colaboración internacional modesta (14,7 %). El mapa temático reveló que los temas motores se centran en el uso genérico de los LLM y de ChatGPT, mientras que las aplicaciones metodológicas específicas aún no conforman agrupaciones consolidadas. **Conclusión:** El uso de la GenAI y de los LLM como apoyo metodológico a la investigación en salud constituye un campo emergente y en rápida expansión, pero que todavía carece de una estructura temática y de canales de publicación consolidados. La consolidación del área exige establecer criterios normativos de validación, transparencia en la comunicación de los resultados, una supervisión humana estricta y la alfabetización de los investigadores en inteligencia artificial, de modo que se garantice la integridad de la evidencia científica producida, con repercusiones directas en la calidad de las prácticas y en la seguridad de la atención en salud.

Descriptores: Bibliometría; Inteligencia Artificial Generativa; Modelos de Lenguaje de Gran Escala; Salud; Técnicas de Investigación.

Editores:

Rosimere Ferreira Santana (ORCID: 0000-0002-4593-3715)
Audrey Vidal Pereira (ORCID: 0000-0002-6570-9016)

Editora:

Escola de Enfermagem Aurora de Afonso Costa – UFF
Rua Dr. Celestino, 74 – Centro, CEP: 24020-091 – Niterói, RJ, Brasil
Correo electrónico de la revista: objn.cme@id.uff.br

Autor correspondiente:

Yasmin Saba de Almeida
E-mail: yasminsaba@id.uff.br

Lo que ya se sabe:

- La inteligencia artificial generativa y los modelos de lenguaje de gran tamaño han ampliado el uso de la inteligencia artificial en actividades de creación, análisis y organización del conocimiento.
- En el ámbito de la salud, estas tecnologías se utilizan para apoyar el diagnóstico, la educación, la redacción académica y el descubrimiento y la síntesis de la literatura científica.
- El uso de la inteligencia artificial generativa como apoyo metodológico en las etapas de la investigación en salud todavía ha sido poco explorado de forma sistemática.

Lo que este artículo aporta:

- Presenta un panorama bibliométrico específico del uso de la inteligencia artificial generativa como apoyo metodológico a la investigación en salud.
- Identifica a los principales autores, revistas, países, redes de colaboración y tendencias temáticas de este campo.
- Evidencia un campo reciente y en rápida expansión e indica prioridades para su consolidación científica y metodológica.

INTRODUCCIÓN

La inteligencia artificial generativa (GenAI), en especial los modelos de lenguaje de gran tamaño (LLM), ha pasado rápidamente de ser un recurso de redacción asistida a aplicarse en dominios de alta complejidad epistémica, incluidas diversas etapas estructurales del proceso de investigación científica⁽¹⁾. En el ecosistema biomédico, este cambio de paradigma ha ido acompañado de la expansión de herramientas asistidas por inteligencia artificial (IA) orientadas al descubrimiento y la exploración de la literatura, que abarcan desde plataformas de minería de textos hasta la representación relacional y semántica del conocimiento^(2,3).

En el ámbito de la investigación en salud, la síntesis de evidencia se ha consolidado como uno de los principales dominios de aplicación de la GenAI y de los LLM. Estos modelos se han ido incorporando progresivamente a la formulación de estrategias de búsqueda, al cribado automatizado o semiautomatizado de la literatura, a la extracción parametrizada de datos y a la evaluación de la calidad metodológica o del riesgo de sesgo. Sin embargo, la evidencia sintetizada en una revisión de alcance reciente advierte que la madurez tecnológica de estas herramientas aún no garantiza su reproducibilidad ni una precisión que prescinda de la supervisión crítica⁽⁴⁾.

Aunque el volumen de publicaciones sobre IA en salud ha crecido de forma exponencial, las investigaciones dedicadas a las aplicaciones estrictamente metodológicas dentro del flujo de investigación siguen diluidas entre estudios centrados en el apoyo al diagnóstico, la educación en salud y el apoyo a la escritura académica. Búsquedas preliminares en PubMed, Scopus y Web of Science confirmaron la ausencia de estudios bibliométricos que mapeen, con rigor científico y de forma sistemática y detallada, el uso de estas tecnologías como apoyo metodológico en la investigación en salud. Este vacío es relevante porque la investigación en salud depende de la producción oportuna y rigurosa de evidencia para orientar la atención, la elaboración de guías y las políticas de salud. Al apoyar etapas que exigen un gran volumen de trabajo, como la búsqueda, el cribado, la extracción y la síntesis de datos, la GenAI y los LLM tienen un elevado potencial para optimizar el ciclo de producción de evidencia, siempre que se garanticen debidamente su validación técnica, su precisión y su reproducibilidad procedimental^(4,5).

En este sentido, comprender cómo se ha ido confi-

gurando este campo permite reconocer qué etapas de la investigación en salud concentran ya los esfuerzos, qué autores, instituciones y fuentes impulsan la producción científica y qué vacíos persisten. Esta visión puede orientar las prioridades de investigación y favorecer un uso de la GenAI y de los LLM más coherente con las necesidades y el rigor de la investigación en salud. El análisis bibliométrico permite examinar el crecimiento de la producción, las redes de colaboración y los temas que estructuran el conocimiento producido, ofreciendo un panorama preciso del grado de madurez de este campo^(6,7). Así, el objetivo de este estudio es mapear la producción científica y las tendencias temáticas del uso de la GenAI y de los LLM como herramientas de apoyo metodológico en la investigación en salud.

MÉTODO

Se trata de un estudio cuantitativo y bibliométrico, caracterizado por la aplicación de métodos cuantitativos para analizar la producción científica en un campo específico del conocimiento. La bibliometría utiliza indicadores estadísticos para examinar patrones de publicación, citación y colaboración científica, lo que permite identificar estructuras del conocimiento, la dinámica de productividad y los dominios temáticos de una disciplina^(6,7).

Pregunta de investigación

La pregunta orientadora del estudio fue: ¿cuál es el perfil bibliométrico y cuáles son las tendencias temáticas de la producción científica sobre la GenAI y los LLM utilizados como herramientas de apoyo metodológico en la investigación en salud?

Criterios de elegibilidad

Se incluyeron documentos en los que la GenAI, especialmente los LLM, se empleara, evaluara o discutiera específicamente como apoyo a al menos una etapa metodológica de la investigación en salud. Se consideraron etapas como la concepción y el diseño metodológico, la elaboración de estrategias de búsqueda, el cribado de títulos, resúmenes o textos completos, la extracción y síntesis de datos, la evaluación de la calidad metodológica o del riesgo de sesgo, el análisis cuantitativo o cualitativo de datos y la generación de hipótesis de investigación.

Se excluyeron los documentos que abordaban estas

tecnologías exclusivamente en contextos de práctica clínica o asistencial, educación o capacitación en salud, redacción científica o edición de manuscritos, así como aquellos que solo mencionaban la IA sin una aplicación metodológica identificable. También se excluyeron los estudios ajenos al campo de la salud.

No se aplicaron restricciones de idioma, tipo de documento ni período de publicación.

Búsqueda bibliográfica

Las búsquedas se realizaron el 1 de septiembre de 2026 en las bases de datos indexadas PubMed, Scopus y Web of Science. Las estrategias combinaron el vocabulario controlado de los *Medical Subject Headings* (MeSH) con términos libres relacionados con la GenAI y los LLM, con el proceso de investigación y con la investigación en salud. También se utilizaron operadores booleanos y términos de exclusión para reducir la recuperación de estudios centrados exclusivamente en aplicaciones clínicas, educación en salud o redacción científica. Las estrategias de búsqueda completas están disponibles para su auditoría en el repositorio del proyecto en Open Science Framework (OSF), con el DOI: 10.17605/OSF.IO/NC8K9.

Selección de estudios asistida por GenAI

Se desarrolló en Python un flujo computacional para automatizar y documentar la deduplicación, la clasificación de los registros y el registro de las decisiones de cribado. El flujo se construyó con el apoyo de las herramientas de GenAI Codex, de OpenAI, y Claude Code, de Anthropic, utilizando Visual Studio Code, versión 1.137.0, como entorno de desarrollo. Los criterios de elegibilidad y el *prompt* estructurado fueron definidos y refinados de forma iterativa por la investigadora a lo largo de este proceso. El código fuente y la versión del *prompt* utilizada en la clasificación están disponibles en el repositorio público de GitHub en [fermagempesquisadora/genai-health-research-screening](https://github.com/fermagempesquisadora/genai-health-research-screening).

Tras la deduplicación por coincidencia exacta y semántica, los títulos y resúmenes se procesaron con el apoyo del LLM Claude. Para cada registro, el modelo indicó la decisión de incluirlo, excluirlo o clasificarlo como incierto, y registró además la categoría de exclusión, cuando correspondía, una justificación concisa, un fragmento del título o del resumen que respaldara la decisión y el nivel de confianza declarado por el propio modelo. Los fragmentos de respaldo se verificaron automáticamente frente al texto fuente. La investigadora realizó una revisión general de las clasificaciones y evaluó de forma prioritaria los registros clasificados como inciertos, definiendo las decisiones finales de cribado bajo supervisión humana integral.

Procesamiento y análisis de los datos

Los datos se extrajeron, armonizaron y analizaron en el entorno R (mediante la interfaz RStudio), con el paquete *bibliometrix*⁽⁶⁾ y su interfaz gráfica, *biblioshiny*. *Bibliometrix* es una herramienta de código abierto orientada al análisis bibliométrico, con recursos avanzados para evaluar la productividad, la colaboración, la citación y el mapeo científico⁽⁶⁾. Se realizaron análisis descriptivos complemen-

tarios en hojas de cálculo de Microsoft Excel.

El análisis informétrico se estructuró a partir de las tres leyes clásicas de la bibliometría: 1) la Ley de Bradford (que evaluó la dispersión de los artículos entre las fuentes de publicación, lo que permitió delimitar empíricamente el núcleo de revistas del corpus)⁽⁸⁾; 2) la Ley de Lotka (que examinó la distribución de la productividad de los autores y puso a prueba el ajuste estadístico entre la distribución observada y el modelo teórico de potencia mediante la prueba de Kolmogórov-Smirnov)⁽⁹⁾; y 3) la Ley de Zipf (que describió la distribución de frecuencias de los términos y sirvió de base para los análisis de coocurrencia de palabras clave y para la estructura conceptual del campo)⁽¹⁰⁾.

También se analizaron el volumen anual de producción, la tasa de crecimiento anual, el tipo y el idioma de los documentos, la antigüedad media de las publicaciones, la distribución geográfica de las afiliaciones de los autores, las instituciones más representadas, los países más productivos y citados, las fuentes y los autores más productivos o citados, la coocurrencia de palabras clave y el mapeo temático. Los indicadores de producción e impacto incluyeron el número de publicaciones, el total de citas y las citas por documento. Las citas se consideraron según los valores presentes en los metadatos de los registros importados y se interpretaron como medidas descriptivas del corpus, teniendo en cuenta las diferencias de cobertura entre las bases de datos⁽¹¹⁾.

El diagrama estratégico situó las agrupaciones temáticas en un espacio bidimensional según la centralidad (grado de relevancia y de conectividad entre grupos) y la densidad (grado de desarrollo interno de la agrupación). Los temas se clasificaron en cuatro cuadrantes: motores, de nicho, emergentes o en declive, y básicos o transversales.

Aspectos éticos

Al utilizar exclusivamente registros bibliográficos e información de acceso público, sin datos identificables de participantes, el estudio no requirió la evaluación del Sistema CEP/Conep, de conformidad con la Resolución n.º 510/2016 del Consejo Nacional de Salud (CNS) de Brasil⁽¹²⁾. Se utilizaron herramientas de GenAI como apoyo al desarrollo del flujo de análisis, al cribado de los registros y a la redacción del manuscrito, bajo la supervisión de la investigadora y la responsabilidad intelectual directa de las autoras.

RESULTADOS

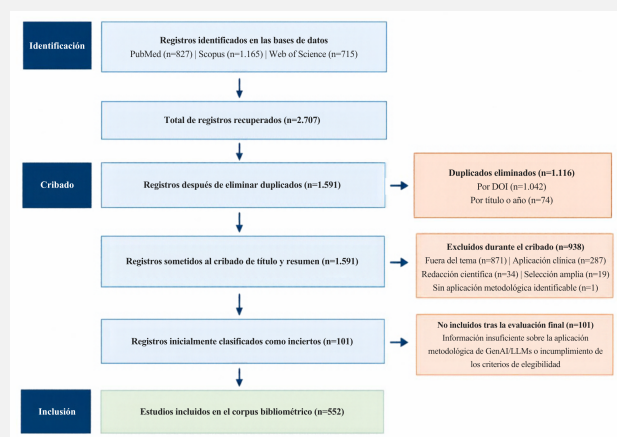
Flujo de selección y elegibilidad del corpus

Se identificaron 2707 registros en las bases de datos seleccionadas. Tras eliminar 1116 duplicados (1042 identificados por DOI y 74 por título y año de publicación), se sometieron 1591 registros únicos al cribado de títulos y resúmenes. El flujo híbrido de cribado (humano-IA) dio como resultado 552 registros elegibles para el corpus bibliométrico y 938 excluidos. De los registros excluidos, 587 no se ajustaban al tema, 287 abordaban aplicaciones clínicas, 34 trataban sobre redacción científica, 29 sobre educación en salud y uno no presentaba una aplicación

metodológica propia identificable.

Los 101 registros clasificados inicialmente como inciertos se excluyeron del corpus final tras una evaluación cualitativa, ya que no fue posible confirmar que la GenAI o los LLM se utilizaran, se evaluaran o sirvieran de apoyo a una etapa metodológica de la investigación en salud. Entre ellos había también estudios ajenos al tema o centrados en aplicaciones clínicas, redacción científica o educación en salud, así como documentos sin contenido metodológico propio. La Figura 1 presenta el flujo de selección de los estudios.

Figura 1 – Diagrama de flujo de la selección de los estudios. Niterói, RJ, Brasil, 2026



Fuente: elaborado por las autoras, 2026.

Características y evolución de la producción científica

Los 552 documentos se distribuyeron en 337 fuentes. Predominaron los artículos originales de revistas (n=425; 77,0 % del corpus), seguidos de las comunicaciones en actas de congresos (n=68), las revisiones (n=38), las cartas (n=13), los editoriales (n=3), los capítulos de libro (n=3) y las notas (n=2). El inglés fue el idioma predominante, presente en el 98,7 % de los registros. La media de citas por documento fue de 2,62.

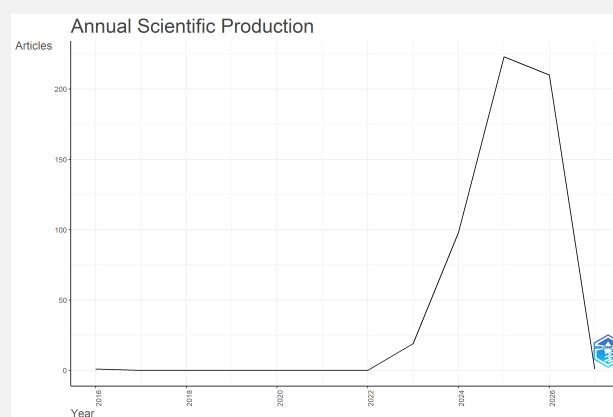
La producción científica abarcó el período bibliográfico de 2016 a 2027 (Figura 2). Se identificó un documento en 2016, sin registros entre 2017 y 2022. A partir de 2023 se observó un aumento notable de la producción, con 19 documentos en 2023, 98 en 2024, 223 en 2025 y 210 hasta el 1 de septiembre de 2026. El único documento registrado en 2027 correspondía a una comunicación de congreso ya disponible en línea en 2026, pero vinculada a un volumen bibliográfico de 2027. El registro se mantuvo en el corpus y no se interpretó como una tendencia de producción posterior al cierre de las búsquedas. La antigüedad media de los documentos fue de 0,88 años, lo que pone de manifiesto la extrema actualidad del campo.

Dispersión de las fuentes de publicación e impacto de citación (Ley de Bradford)

La distribución de las 337 fuentes en zonas de Bradford identificó un núcleo de 27 fuentes, responsable de 187 documentos, es decir, el 33,9 % del corpus (Figura 3). La segunda zona reunió 127 fuentes y 182 documentos, mien-

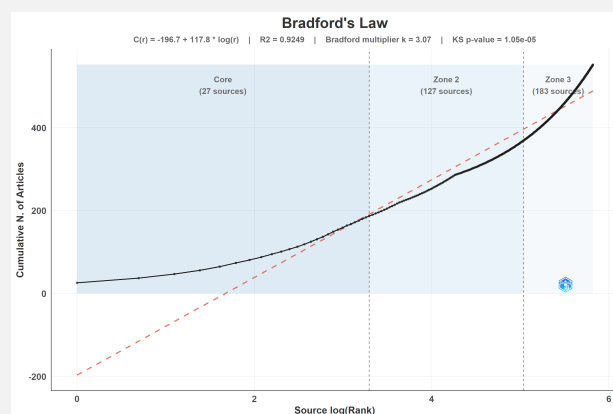
tras que la tercera incluyó 183 fuentes y 183 documentos. El modelo de Bradford mostró un excelente ajuste lineal ($R^2 = 0,9249$), con un multiplicador de 3,07. La prueba de Kolmogórov-Smirnov confirmó la discontinuidad estadística entre la distribución empírica observada y el modelo teórico ideal ($D = 0,1899$; $p < 0,001$).

Figura 2 – Producción científica anual sobre la GenAI y los LLM como herramientas de apoyo metodológico en la investigación en salud, 2016–2027 (n=552). Niterói, RJ, Brasil, 2026



Nota: El único documento con año bibliográfico 2027 estaba disponible en línea en 2026 y se mantuvo en el corpus por haber sido recuperado durante el período de búsqueda. Fuente: datos de la investigación, generados en Biblioshiny, 2026.

Figura 3 – Distribución de las fuentes en zonas de Bradford. Niterói, RJ, Brasil, 2026



Fuente: datos de la investigación, generados en Biblioshiny, 2026.

La revista *Studies in Health Technology and Informatics* se destacó como la fuente más productiva (n=26). En cuanto al impacto de citación, el *Journal of Clinical Epidemiology* encabezó el corpus, con 149 citas acumuladas en seis publicaciones (Tabla 1).

Entre los artículos con mayor número global de citas, destacaron los dedicados a la exploración semántica y la síntesis de evidencia. El artículo de Wei CH, publicado en 2024 sobre la plataforma PubTator 3.0, fue el más citado del corpus (n=104), seguido de los estudios de Lieberum JL (2025; n=103), Aljamaan F (2024; n=97), Wang S (2024; n=92) y Jin Q (2024; n=91).

Tabla 1 – Principales fuentes del corpus según el número de publicaciones y las citas disponibles. Niterói, RJ, Brasil, 2026

Fuente	NP	TC	CPP (TC/NP)	% del total (NP/552 × 100)
Studies in Health Technology and Informatics	26	11	0,42	4,71
Journal of Medical Internet Research	11	0	0,00	1,99
bioRxiv: The Preprint Server for Biology	10	0	0,00	1,81
JMIR AI	9	0	0,00	1,63
Journal of the American Medical Informatics Association	9	0	0,00	1,63
Scientific Reports	9	0	0,00	1,63
Bioinformatics (Oxford, England)	7	0	0,00	1,27
BMC Medical Research Methodology	7	33	4,71	1,27
Cureus	7	0	0,00	1,27
BMJ Open	6	1	0,17	1,09

Nota: NP = número de publicaciones; TC = total de citas; CPP = citas por publicación.

Fuente: datos de la investigación, 2026.

Patrones de autoría, productividad y cooperación (Ley de Lotka)

El corpus reunió a 2914 autores. Wang Y fue el autor más productivo (n=18), seguido de Wang Z (n=13), Chen Y (n=12) y Li Y (n=12). La producción fue predominantemente colaborativa, con una media de 6,32 coautores por artículo. Solo 21 documentos, que corresponden al 3,8 %, se publicaron con autoría individual. El índice de colaboración internacional fue del 14,7 %.

La aplicación de la Ley de Lotka reveló que 2599

autores (el 89,2 % del total) publicaron un solo artículo en el área (Tabla 2). La distribución empírica difirió del modelo teórico clásico de la Ley de Lotka, con un exponente $\beta = 2,0$ ($D = 0,6429$; $p = 0,006$). No obstante, la estimación del modelo de potencia ajustado al corpus arrojó $\beta = 2,87$ (constante $C = 0,444$ y $R^2 = 0,928$), sin diferencias estadísticamente significativas con respecto a los datos observados ($D = 0,2857$; $p = 0,617$). Este elevado exponente β refleja una concentración atípica de autores ocasionales, característica de los campos emergentes.

Tabla 2 – Distribución de los autores según el número de documentos publicados y proporción teórica de la Ley de Lotka clásica. Niterói, RJ, Brasil, 2026

Número de publicaciones	Número de autores	Proporción observada (%)	Proporción teórica (%)
1	2599	89,2	60,8
2	210	7,2	15,2
3	60	2,1	6,8
4	15	0,5	3,8
5	10	0,3	2,4
6	5	0,2	1,7
7	5	0,2	1,2
8	1	0,0	0,9
9	1	0,0	0,8
10	3	0,1	0,6
11	1	0,0	0,5
12	2	0,1	0,4
13	1	0,0	0,4
18	1	0,0	0,2

Nota: La proporción teórica corresponde al modelo clásico de la Ley de Lotka, calculado mediante $f(n) = 0,6079/n^2$.

Fuente: datos de la investigación, 2026.

Desde el punto de vista geográfico, Estados Unidos encabezó la frecuencia de afiliaciones (881 apariciones), seguido de China (502 apariciones). La institución más frecuente fue la Lanzhou University (n=102), seguida del National Institute on Aging (n=47), la University of California (n=44), la McMaster University (n=37) y la University of Pennsylvania (n=35). En cuanto a las citas por país de afiliación, Estados Unidos acumuló 228 citas (media de 9,9 citas por artículo), mientras que China sumó 14 citas (media de 2,0 por artículo). Jin Q obtuvo el mayor número de citas entre los autores, con 195, mientras que Wang Y se destacó por su productividad.

Estructura conceptual y mapa temático

El mapa temático bidimensional (Figura 4), construido a partir de las palabras clave de los autores, organizó la

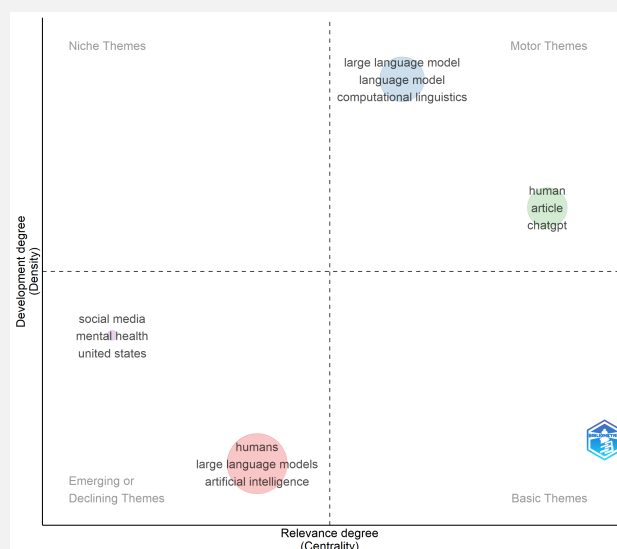
estructura conceptual en cuatro agrupaciones, a saber:

Temas motores (cuadrante superior derecho): se identificaron dos clústeres bien definidos. El primero agrupó los términos “*large language model*”, “*language model*” y “*computational linguistics*”; el segundo reunió “*ChatGPT*”, “*human*” y “*article*”, lo que indica el fuerte desarrollo y la centralidad metodológica de los LLM en el ecosistema de investigación.

Temas emergentes o en declive (cuadrante inferior izquierdo): se agruparon los términos “*humans*”, “*large language models*” y “*artificial intelligence*”, además del clúster formado por “*social media*”, “*mental health*” y “*United States*”.

Temas de nicho y básicos: no se identificaron agrupaciones estructuradas en estos cuadrantes, lo que refuerza el carácter concentrado y todavía dinámico de la literatura analizada.

Figura 4 – Mapa temático de las palabras clave de los autores asociadas a la GenAI y a los LLM como herramientas de apoyo metodológico en la investigación en salud. Niterói, RJ, Brasil, 2026



Fuente: datos de la investigación, generados en Biblioshiny, 2026.

DISCUSIÓN

Dinámica de crecimiento y grado de consolidación del campo

En comparación con estudios bibliométricos más amplios sobre la GenAI en salud, la producción centrada estrictamente en el apoyo metodológico a la investigación es todavía más acotada y se encuentra en una fase inicial de consolidación. Un estudio bibliométrico sobre ChatGPT en salud, que incluyó aplicaciones clínicas, educativas y de comunicación con pacientes, reunió 594 publicaciones y registró un crecimiento anual del 91,18 % entre 2023 y 2024⁽¹³⁾. Otro análisis, centrado en el uso de ese mismo LLM en la educación en Asia, identificó 1665 documentos publicados entre 2022 y 2025, con una tasa de crecimiento anual del 545,53 %⁽¹⁴⁾. En el presente estudio, esta tasa fue del 70,7 %, inferior a las observadas en ambos análisis. Las diferencias de bases de datos, períodos y estrategias de búsqueda limitan una comparación directa entre estos resultados. Aun así, los hallazgos muestran que el uso de la GenAI y de los LLM como apoyo metodológico a la investigación en salud también está en expansión, aunque de forma menos acentuada y con mayor rigor en la selección de los casos de uso.

Informetría

de las redes de autoría, las fuentes y la cooperación transnacional

El análisis de las leyes de Bradford y de Lotka, junto con los indicadores de colaboración, apunta en la misma dirección. Aunque algunas revistas ya concentran una parte relevante de las publicaciones, los documentos se distribuyen entre 337 fuentes, lo que indica que el tema aún no cuenta con medios de difusión plenamente consolidados.

La productividad de los autores también se mostró fuertemente concentrada en contribuciones puntuales: el 89,2 % de ellos aparece en un solo documento, lo que configura una elevada proporción de autores transitorios, y la caída de la productividad es más pronunciada que la prevista por el modelo clásico de Lotka ($\beta = 2,87$). La colaboración internacional, presente en el 14,7 % de los documentos, completa este panorama al mostrar que las conexiones entre países siguen siendo poco frecuentes. En conjunto, estos resultados sugieren que el crecimiento de la literatura se ha producido más por la entrada de nuevos autores y por la dispersión del tema entre distintas revistas que por la formación de una comunidad de investigación estable y consolidada.

Hegemonía geográfica e implicaciones para la equidad científica

La concentración de las afiliaciones en Estados Unidos y China debe interpretarse a la luz de las diferencias entre países en cuanto a infraestructura, gobernanza, fuerza laboral y capacidad institucional para la adopción de la IA en salud⁽¹⁵⁾. Aunque este estudio no evaluó directamente estas desigualdades, los resultados muestran que una parte relevante del conocimiento sobre el uso de la GenAI en la investigación en salud se está produciendo en países con una elevada capacidad científica y tecnológica. Este predominio puede influir en las aplicaciones investigadas, en los criterios utilizados para evaluarlas y en los problemas considerados prioritarios. Por lo tanto, las tecnologías desarrolladas y validadas predominantemente en estos contextos podrían no responder de la misma manera a las necesidades de los sistemas de salud y de las poblaciones que operan en condiciones de escasez de recursos o con particularidades epidemiológicas distintas.

Evolución de la estructura y madurez temática

El mapa temático mostró que los temas más desarrollados y conectados siguen centrados en la evaluación genérica de los LLM, con ChatGPT en un lugar destacado. Esto indica que la literatura todavía se organiza en torno a las propias tecnologías y a sus posibilidades de uso, mientras que las aplicaciones metodológicas más específicas (como la automatización de síntesis de evidencia o la modelización del riesgo de sesgo) aún no aparecen como temas consolidados. La ausencia de temas básicos o de nicho y la clasificación de las demás agrupaciones como emergentes o en declive refuerzan este panorama. Así, aunque la GenAI ya se está incorporando a distintas etapas de la investigación en salud, todavía no se observa una estructura temática bien definida en torno a usos metodológicos específicos.

Incertidumbres metodológicas, integridad y reproducibilidad

La literatura ya describe el uso de la GenAI en la búsqueda, el cribado, la extracción de datos, la evaluación del riesgo de sesgo y la síntesis de evidencia. Sin embargo, todavía faltan criterios estandarizados y consistentes que definan en qué etapas de la investigación pueden utilizarse los LLM con seguridad, qué niveles de error son aceptables

y cómo debe auditarse y validarse su desempeño^(4,5). En la investigación en salud, esta incertidumbre es especialmente relevante, ya que los errores en la identificación, selección o interpretación de la evidencia pueden comprometer revisiones que sustentan protocolos, guías, políticas y decisiones asistenciales, con posibles repercusiones en la calidad y la seguridad de la atención⁽⁵⁾.

Por ello, la incorporación de la GenAI a la investigación en salud no debe guiarse solo por su potencial para agilizar etapas del trabajo, sino también por su capacidad de preservar el rigor, la fiabilidad y la reproducibilidad de la evidencia producida. La comunicación de su uso debe incluir la herramienta y su versión, la finalidad de uso, las instrucciones empleadas, las formas de validación, la verificación de los resultados y los aspectos de privacidad implicados⁽¹⁶⁾. Este nivel de detalle ayuda a comprender los beneficios y las limitaciones de la tecnología en el contexto en que se utilizó y permite auditar las decisiones tomadas a lo largo del proceso. Aun con el uso de estas herramientas, la planificación, la validación y la interpretación de los resultados siguen siendo responsabilidad exclusiva de los investigadores⁽¹⁶⁾.

Alfabetización en IA y transformación de la identidad profesional del investigador

Por último, la expansión del uso de la GenAI y de los LLM en el proceso de investigación plantea nuevos desafíos para la formación de los investigadores. Su uso requiere no solo el dominio operativo de las herramientas, sino también alfabetización en IA y la capacidad de evaluar críticamente los resultados generados, reconocer las limitaciones algorítmicas y justificar las decisiones metodológicas. Una revisión sistemática sobre educación y capacitación en IA para trabajadores de la salud señaló la necesidad de iniciativas formativas que contemplen competencias técnicas, evaluación crítica, aspectos éticos y un uso seguro de estas tecnologías⁽¹⁷⁾, mientras que otro estudio mostró que la identidad profesional puede estar relacionada con la preparación para el uso de la IA entre los trabajadores de la salud⁽¹⁸⁾. Estos hallazgos sugieren que la formación de los investigadores debe acompañar estos cambios, considerando no solo las competencias necesarias para el uso de la GenAI, sino también la comprensión de su papel y de su responsabilidad ante las nuevas formas de producir conocimiento^(17,18).

Limitaciones del estudio

Como limitaciones, los resultados reflejan las bases de datos consultadas, la estrategia de búsqueda y el período adoptado. Los indicadores de citación se obtuvieron de los metadatos importados y deben interpretarse como medidas descriptivas del corpus, ya que la cobertura varía entre las fuentes. También debe considerarse el uso de un LLM como apoyo al cribado, dado que las clasificaciones automatizadas están sujetas a errores y pueden haber influido en la composición del corpus, pese a la supervisión humana. Al mismo tiempo, su uso en el propio proceso de selección permitió explorar, en la práctica, una de las aplicaciones metodológicas que constituyen el objeto de este

estudio. Además, como la producción sobre el tema crece rápidamente, el panorama presentado está condicionado al período analizado y puede modificarse con la publicación de nuevos estudios.

CONCLUSIÓN

Este estudio mostró que el uso de la GenAI y de los LLM como apoyo metodológico a la investigación en salud constituye un campo reciente y en expansión, pero todavía marcado por la dispersión de las publicaciones entre autores y revistas, una baja colaboración internacional y la ausencia de una estructura temática consolidada en torno a aplicaciones metodológicas específicas. Los hallazgos indican que la incorporación de estas tecnologías avanza más rápido que la construcción de una base científica y regulatoria capaz de orientar sus usos en la investigación en salud.

Se recomienda que los estudios futuros validen el desempeño de estas herramientas en etapas metodológicas específicas, que su uso se comunique de forma estandarizada y transparente, con supervisión humana, y que la alfabetización en IA se incorpore a la formación de los investigadores, también en enfermería. Se recomiendan, asimismo, el fortalecimiento de las redes colaborativas, la ampliación de los estudios en países del Sur Global, que tienden a incorporar tecnologías de IA desarrolladas en otros contextos, y la actualización periódica de este mapeo. Estas medidas son necesarias porque los errores en la búsqueda, selección, análisis y síntesis de la evidencia pueden repercutir en la calidad y la seguridad de la atención, así como en las guías y políticas de salud.

CONFLICTO DE INTERESES

Los autores declaran no tener conflicto de intereses.

FINANCIACIÓN

El presente trabajo fue realizado con el apoyo de la Coordinación de Perfeccionamiento de Personal de Nivel Superior – Brasil (CAPES) – Código de Financiamiento 001.

DECLARACIÓN DE USO DE INTELIGENCIA ARTIFICIAL

Se utilizaron herramientas de GenAI en distintas etapas de este estudio. En el desarrollo del flujo computacional en Python se utilizaron Codex y Claude Code. El cribado de títulos y resúmenes se realizó con el apoyo de Claude Opus 5, a partir de criterios de elegibilidad y de un prompt definidos y refinados por la investigadora. ChatGPT y Claude también se utilizaron como apoyo para la organización, la redacción y la revisión lingüística del manuscrito, así como para la elaboración visual del diagrama de flujo de selección de los estudios.

El uso de estas herramientas se realizó bajo supervisión humana integral. Las clasificaciones generadas por Claude durante el cribado fueron revisadas por la investigadora, que evaluó de forma prioritaria los registros clasificados como inciertos y ajustó las decisiones cuando fue necesario.

La definición de los criterios de elegibilidad, las decisiones metodológicas, la interpretación de los resultados y las conclusiones fueron responsabilidad de las investigadoras,

que revisaron críticamente la versión final y asumen la plena responsabilidad por el contenido presentado.

REFERENCIAS

1. Truhn D, Reis-Filho JS, Kather JN. Large language models should be used as scientific reasoning engines, not knowledge databases. *Nat Med*. 2023;29(12):2983-2984. <https://doi.org/10.1038/s41591-023-02594-z>. PMID: 37853138.
2. Wei C, Allot A, Lai P, Leaman R, Tian S, Luo L, et al. PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge. *Nucleic Acids Research*. 2024;52(W1):W540-W546. <https://doi.org/10.1093/nar/gkae235>. PMID: 38572754.
3. Jin Q, Leaman R, Lu Z. PubMed and beyond: biomedical literature search in the age of artificial intelligence. *eBioMedicine*. 2024;100:104988. <https://doi.org/10.1016/j.ebiom.2024.104988>. PMID: 38306900.
4. Lieberum J, Toews M, Metzendorf M, Heilmeyer F, Siemens W, Haverkamp C, et al. Large language models for conducting systematic reviews: on the rise, but not yet ready for use—a scoping review. *Journal of Clinical Epidemiology*. 2025;181:111746. <https://doi.org/10.1016/j.jclinepi.2025.111746>. PMID: 40021099.
5. Siemens W, von Elm E, Binder H, Böhringer D, Eisele-Metzer A, Gartlehner G, et al. Opportunities, challenges and risks of using artificial intelligence for evidence synthesis. *BMJ EBM*. 2025;30(6):381-384. <https://doi.org/10.1136/bmjebm-2024-113320>. PMID: 39788693.
6. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*. 2017;11(4):959-975. <https://doi.org/10.1016/j.joi.2017.08.007>.
7. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*. 2021;133:285-296. <https://doi.org/10.1016/j.jbusres.2021.04.070>.
8. Bradford SC. Sources of information on specific subjects. *Engineering*. 1934;137:85-86.
9. Lotka AJ. The frequency distribution of scientific productivity. *J Wash Acad Sci*. [Internet]. 1926 [citado 12 sep 2026];16(12):317-23. Disponible en: <https://www.jstor.org/stable/24529203>.
10. Zipf GK. Human behavior and the principle of least effort: an introduction to human ecology. Cambridge: Addison-Wesley; 1949.
11. Gerasimov I, KC B, Mehrabian A, Acker J, McGuire MP. Comparison of datasets citation coverage in Google Scholar, Web of Science, Scopus, Crossref, and DataCite. *Scientometrics*. 2024;129(7):3681-3704. <https://doi.org/10.1007/s11192-024-05073-5>.
12. Brasil. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016 [Resolución n.º 510, de 7 de abril de 2016. Establece las normas aplicables a las investigaciones en Ciencias Humanas y Sociales cuyos procedimientos metodológicos impliquen el uso de datos obtenidos directamente de los participantes o de información identificable, o que puedan conllevar riesgos mayores que los de la vida cotidiana, en la forma definida en esta Resolución] [Internet]. Brasília: Ministério da Saúde; 2016 [citado 12 sep 2026]. Disponible en: <https://www.gov.br/conselho-nacional-de-saude/pt-br/atos-normativos/resolucoes/2016/resolucao-no-510.pdf/view>.
13. Amir AS. Revolutionizing healthcare with ChatGPT: current advancements, challenges, and future directions. *Discov Artif Intell*. 2026;6(1):582. <https://doi.org/10.1007/s44163-026-01506-7>.
14. Irwanto I. Research on ChatGPT in education in Asia: a bibliometric analysis (2022–2025). *Discov Artif Intell*. 2025;5(1):244. <https://doi.org/10.1007/s44163-025-00505-4>.
15. World Health Organization, Regional Office for Europe. Artificial intelligence is reshaping health systems: state of readiness across the European Union [Internet]. Copenhagen: WHO Regional Office for Europe; 2026 [citado 12 sep 2026]. Disponible en: <https://www.who.int/europe/publications/i/item/WHO-EURO-2026-12707-52481-81471>.
16. 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 EBM*. 2025;30(6):390-400. <https://doi.org/10.1136/bmjebm-2025-113825>. PMID: 40360239.
17. Woods L, Lyons K, Van Der Vegt A, Olsen Q, Huang W, Khor JS, et al. Assessing the effectiveness of artificial intelligence education and training for healthcare workers: a systematic review. *BMC Med Educ*. 2026;26(1). <https://doi.org/10.1186/s12909-026-08969-3>. PMID: 41807961.
18. Ta'an W, Damrah S, Al-Hammouri MM, Williams B. Professional identity and its relationships with AI readiness and inter-professional collaboration. *PLoS One*. 2025;20(5):e0322794. <https://doi.org/10.1371/journal.pone.0322794>. PMID: 40378129.

CONTRIBUCIÓN DE LOS AUTORES

Diseño del proyecto: Almeida YS, Cortez EA, Valente GSC.

Obtención de datos: Almeida YS.

Análisis de datos: Almeida YS.

Interpretación de datos: Almeida YS.

Todos los autores asumen la responsabilidad por la redacción del manuscrito, la revisión crítica de su contenido intelectual, la aprobación de la versión final publicada y todos los aspectos éticos, legales y científicos relacionados con la exactitud y la integridad del estudio.



Copyright © 2026 Online Brazilian Journal of Nursing

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.