

El desafío de la inteligencia artificial generativa en la publicación científica: consideraciones para un proceso editorial ético

The challenge of generative artificial intelligence in scientific publishing: considerations for an ethical editorial process

Correspondencia

Yolanda Angulo-Bazán
yangulob@cientifica.edu.pe

Recibido: 18/09/2024

Aprobado: 25/09/2024

Citar como: Angulo-Bazán Y. El desafío de la inteligencia artificial generativa en la publicación científica: Consideraciones para un proceso editorial ético. *Acta Med Peru.* 2024;41(3):153-7. doi: 10.35663/amp.2024.413.3328

Este es un artículo Open Access publicado bajo la licencia Creative Commons

Atribución 4.0 Internacional. (CC-BY 4.0)



Yolanda Angulo-Bazán^{1,a,b}

¹ Grupo de Investigación en Bioética e Integridad Científica, Universidad Científica del Sur. Lima, Perú.

^a Médico - cirujano

^b Magister en Epidemiología Clínica

En los últimos años, la inteligencia artificial (IA) generativa se ha involucrado definitivamente y transversalmente en todos los procesos involucrados en la investigación científica en salud, justamente debido a su capacidad de “generar” respuestas concretas en todo tipo de formatos a interrogantes formuladas, denominadas *prompts* ^[1]. Esta situación pone a prueba la capacidad de adaptación y la resolución oportuna de dilemas bioéticos que surgen naturalmente ante la concientización de que el conocimiento generado por IA será aplicado en problemas complejos relacionados a un concepto intrínsecamente complejo, como la salud humana ^[2]. La generación de conocimiento tampoco deja de ser un problema complejo que, por ser un proceso humano, no deja de ser falible, a pesar de los múltiples esfuerzos que los actores involucrados realizan por sistematizar y brindar objetividad a este proceso ^[3].

En ese sentido, la difusión del conocimiento realizada usualmente mediante la publicación en medios como revistas científicas a través de un proceso de revisión por pares, también enfrenta el desafío de la inclusión de la IA generativa, llevando a los comités editoriales a la obligación de informar sobre los potenciales riesgos y beneficios de esta situación y tomar posiciones informadas que salvaguarden la calidad de sus procesos.

A raíz de la popularización del uso *chatbots* como ChatGPT, Microsoft Copilot, Google Gemini, entre otros para redacción científica; organizaciones internacionales como la *Committee on Publication Ethics* (COPE), *World Association of Medical Editors* (WAME) o *International Council of Medical Journal Editors* (ICMJE) se han visto en la necesidad de emitir documentos de consenso enfocados en promover entre los investigadores la práctica de valores como la responsabilidad, ante la imposibilidad de atribuir autoría a una IA, o la transparencia, al reportar de forma honesta todo el proceso de uso de estas tecnologías durante la generación de conocimiento científico ^[4-6].

La interacción de la IA con el proceso de revisión por pares y sus actores (editores y revisores) es integral y presenta numerosas posibilidades ^[7]; no obstante, este punto aún no es discutido en estos consensos, reduciendo la generación de propuestas al llamado a la transparencia en el uso y que los editores puedan contar con herramientas que les ayuden a detectar contenidos generados o modificados por IA en los manuscritos ^[5]. Por ejemplo, uno de los usos de IA más reportados ha sido el de la evaluación de manuscritos que ingresan al proceso editorial, tanto por parte de los editores (a modo de evaluación preliminar) al igual que los revisores (como un complemento a la revisión por pares) ^[8,9].

Una evidencia previa muestra que probablemente el 17% de las revisiones por pares que se desarrollan (por ejemplo, en congresos científicos) podrían estar siendo generadas por herramientas de IA generativas con modelos lingüísticos grandes (LLM, del inglés *large language model*) ^[10]. Así mismo,

se ha evaluado el desempeño de algunas IA generativas como ChatGPT en el papel de revisores, mostrando una coherencia de hasta el 89% comparado con recomendaciones realizadas por revisores humanos ^[11,12]. Sin embargo, la recomendación o implementación del uso de estas tecnologías no solo puede depender de medidas de eficacia ⁽⁷⁾, puesto que aún no pueden reemplazar la experticia, la valoración crítica o la evaluación de la originalidad de un manuscrito que realiza un revisor humano ^[8,13]. Por ende, la carga valorativa y decisiva sobre la publicación o no de un manuscrito y de todo el proceso editorial sigue siendo de responsabilidad de los editores ^[14].

Una de las mayores preocupaciones en estas situaciones, nuevamente gira en torno a la transparencia; por ejemplo, una considerable cantidad de IA generativas se desarrolla bajo un principio de *black-box* (caja negra) en el que los usuarios podemos ver las entradas (*prompts*) y las salidas (respuestas), pero no como éstas se producen o qué información se utiliza para la generación de respuestas que incluyen una valoración ^[15], como en el caso de las revisiones por pares. Este vacío lleva a consecuencias como no poder predecir situaciones en las que hay una mayor probabilidad de error en las conclusiones o incluso de alucinaciones de la IA, que se producen como un intento de obtener respuestas forzadas extrapolando criterios entrenados previamente, pero sin una base real ^[16].

Por otro lado, el manejo ético de la información que se introduce a una IA representa también un dilema bioético que no se restringe a las de tipo generativo. En primer lugar, porque ya se tiene reconocidos, sesgos de información y de selección en los datos que sirven para el entrenamiento de estas tecnologías ^[17], lo que ha llevado a reconocer que la aplicación directa de algoritmos de IA en situaciones humanas puede llevar a situaciones de discriminación, como el llamado “racismo algorítmico” ^[18,19]. Por lo tanto, al no conocer el origen y la naturaleza de los datos que han entrenado a estas tecnologías, el tomar solo en cuenta las recomendaciones finales de una IA generativa en un proceso de revisión por pares puede predisponer a tomar decisión sobre la publicación de un manuscrito sin considerar aspectos de relevancia social o cultural cruciales para la población objetivo, o el alcance de la revista.

Por ejemplo, es importante considerar que los porcentajes de coherencia entre revisores humanos e IAs previamente descritos fueron calculados utilizando información de revistas que publican en idioma inglés y que mayormente no publican evidencia dirigida a Latinoamérica ^[9,11,12,20], básicamente debido a la menor proporción de generación de conocimiento en la región por barreras socioeconómicas ampliamente conocidas ^[21].

Adicionalmente, tampoco se conoce cómo estas IAs gestionan la información que se incluye en ellas, lo que puede ser no solo una barrera legal, por la posibilidad de estar compartiendo información que puede ser confidencial; sino también ética para su uso dentro del proceso de revisión por pares ^[13]. Es importante recordar que el proceso de generación de conocimiento científico y de la difusión del mismo a través de

una revista se basa en el mantenimiento de un valor clave, la confianza ^[22]. Los investigadores deben tener completa confianza en que el manuscrito que envían a una revista será evaluado no solo rigurosamente, sino también, de una manera justa y bajo estándares éticos. Por lo tanto, así como no sería ético fomentar un ambiente donde los autores y editores no sean transparentes con el uso de herramientas de IA, es igualmente importante promover el uso justo de estas tecnologías, lo que implica en este punto establecer un equilibrio entre las necesidades del Comité editor y la capacidad decisional de los autores sobre sus manuscritos.

Por último, es claro que no se está ante un desafío futuro sino ante algo que ya está ocurriendo en la revisión por pares, más aún en un contexto en el que ya se usa en la elaboración de políticas públicas, como las Prioridades Nacionales en Salud ^[23]; y se sugieren tasas altas de uso de *chatbots* en estudiantes de ciencias de la salud ^[24].

En el Perú aún se evidencia una discordancia entre la realidad y la capacidad de respuesta de los actores involucrados, en este caso, las revistas. Por ejemplo, solo dos revistas biomédicas nacionales indexadas en Scopus tienen consideraciones relacionadas al uso de la IA, una de ellas específica para los autores ^[25], alineada a los consensos internacionales previamente mencionados (Tabla 1) y otra que sí amplía la transparencia de uso al comité editorial y a los revisores, y explícita que su uso solo se da cuando es estrictamente necesario ^[26].

En ese sentido, el Comité Editorial de Acta Médica Peruana se encuentra en un proceso de adaptación y actualización de instrucciones para autores donde se incluirán consideraciones específicas sobre el uso de IA generativa; sin embargo, más allá del plano solamente declarativo, es importante asumir posiciones que permitan reforzar la naturaleza humana y la confianza en la revisión por pares a través de principios de transparencia y responsabilidad (Tabla 2), procurando solucionar, además, problemas de larga data en estos procesos, como la escasez de revisores. Es necesario aceptar que todos estamos expuestos al uso de IA y la clave de no caer en procesos de cegamiento valorativo (para bien o para mal) de la misma, es promover un uso responsable y cauteloso.

Empero, esto debe ir de la mano con un pedido de apertura y transparencia hacia las empresas que gerencian estas tecnologías, ya que no se puede gestionar la cautela y los preconceptos en el uso de IA mientras sus creadoras no muestren los mecanismos concretos involucrados en el análisis y desarrollo de valoraciones brindadas por sus tecnologías. Todos los actores involucrados en el desarrollo de conocimiento científico asumen desde un inicio el compromiso de evitar cualquier posibilidad de aprovechamiento indebido y perjuicio a los investigadores y por ende a los sujetos de estudio. La IA y, sobre todo, las empresas detrás de ellas ya son incluidas como parte de estos actores, por lo que las posturas éticas y responsables deben ser no solo una exigencia sino un principio.

Tabla 1. Consideraciones sobre uso de IA generativa en consensos internacionales y políticas de casas editoriales.

| Organización | Título de la declaración* | Año | Recomendación | Fuente |
|--|---|------|--|---|
| Committee of Publication Ethics (COPE) | Authorship and AI tools | 2023 | <ul style="list-style-type: none"> La IA no puede ser autora al no tener responsabilidad de lo que se publica. No es una entidad [con responsabilidad] legal. Se recomienda transparencia en el uso de estas (escritura del manuscrito, producción de imágenes o elementos gráficos, recolección y análisis de datos), cómo fue usada y qué IA fue usada. Los autores son responsables por el contenido del manuscrito, incluso en las partes producidas por IA. | https://publicationethics.org/cope-position-statements/ai-author |
| World Association of Medical Editors (WAME) | Chatbots, Generative AI, and Scholarly Manuscripts WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications | 2023 | <ul style="list-style-type: none"> Los <i>chatbots</i> no pueden ser autores. Los autores deben ser transparentes cuando se utilizan <i>chatbots</i> y proporcionar información sobre cómo se utilizaron. Los autores son responsables del material proporcionado por un <i>chatbot</i> en su artículo. Los editores y revisores deben especificar, a los autores y entre sí, cualquier uso de <i>chatbots</i> en la evaluación del manuscrito y la generación de revisiones y correspondencia. Los editores necesitan herramientas adecuadas que les ayuden a detectar contenidos generados o alterados por IA. | https://wame.org/page3.php?id=106#Spanish |
| International Committee of Medical Journal Editors (ICMJE) | Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Responsibilities in the submission and peer-review process. | 2023 | <ul style="list-style-type: none"> Las revistas deben pedir a los autores que declaren todo uso de IA generativa (especialmente modelos lingüísticos como LLM). Los <i>chatbots</i> no pueden ser autores, no pueden ser responsables de la exactitud, integridad y originalidad del trabajo. Los revisores deben evitar subir contenido de los manuscritos a IAs generativas. Los revisores deben declarar el uso de IA cuidadosamente y ser conscientes de sus desventajas (alucinaciones). El contenido obtenido por IA no es referenciable como primera fuente. | https://www.icmje.org/icmje-recommendations.pdf |

* del inglés *statement*.

IA: Inteligencia artificial; LLM: Modelos lingüísticos grandes.

Tabla 2. Recomendaciones para un manejo ético de la IA generativa dentro del proceso de revisión por pares en revistas biomédicas.

| Principios | Actores involucrados | Recomendación |
|------------------------|--------------------------|---|
| Transparencia | Autores | <ul style="list-style-type: none"> • Declarar detalladamente el uso de herramientas de IA generativa, promoviendo la reproducibilidad de los métodos usados. |
| | Editores | <ul style="list-style-type: none"> • Declarar la posibilidad del uso de herramientas de IA para fines concretos (por ejemplo, evaluación de plagio, evaluación de lenguaje generado por IA o evaluación preliminar de manuscritos en búsqueda de fallas metodológicas) en la sección de política editorial e instrucciones para autores. |
| | Revisores | <ul style="list-style-type: none"> • Declarar detalladamente el uso de herramientas de IA incluyendo el papel que jugó para la toma de la decisión editorial propuesta. |
| | Empresas gestoras de IAs | <ul style="list-style-type: none"> • Visibilizar los métodos que la herramienta utiliza para la toma de decisiones, incluidos criterios de selección de información relevante e información utilizada como sustento. • Visibilizar los mecanismos involucrados en el manejo de la información que se sube a la plataforma. |
| Responsabilidad | Autores | <ul style="list-style-type: none"> • Asumir la responsabilidad de todo lo que se coloca en el manuscrito enviado, independientemente de que el contenido haya sido creado con ayuda de herramientas de IA. • No utilizar herramientas de IA para delegar responsabilidades que son propias de los investigadores (por ejemplo, elaboración de la pregunta de investigación, selección de métodos usados, inclusión o no de datos controversiales, decisión de revista o revistas para publicación, etc.). • Entrenarse adecuadamente en el uso de herramientas de IA generativa, entendiendo sus limitaciones y posibles sesgos. |
| | Editores | <ul style="list-style-type: none"> • Promover el uso de herramientas de IA de libre acceso o con políticas de manejo de datos y metodologías claras. • Brindar herramientas claras y de fácil entendimiento para la declaración de uso de IA en los manuscritos recibidos. • Promover la capacitación de la comunidad científica en uso responsable de herramientas de IA generativa. • Brindar recomendaciones constructivas, intentando eliminar juicios de valor, ante usos inadecuados de IA por parte de autores o revisores. • Entrenarse adecuadamente en el uso de herramientas de IA generativa, entendiendo sus limitaciones, posibles sesgos y funcionamiento para detección de posibles textos fabricados o fraude científico. |
| | Revisores | <ul style="list-style-type: none"> • Realizar un uso responsable de la IA generativa, sin que ésta elimine criterios humanos de valorización de manuscritos (originalidad, revisión crítica, necesidad de publicación en el contexto). • Entrenarse adecuadamente en el uso de herramientas de IA generativa, entendiendo sus limitaciones y posibles sesgos. |
| | Empresas gestoras de IAs | <ul style="list-style-type: none"> • Brindar mensajes de advertencia y recomendación de uso responsable de IA ante preguntas que involucren la revisión de contenidos aún no publicados. • Comprometerse al desarrollo de herramientas éticas y de acceso abierto especializadas en el manejo responsable de la información y usando información de entrenamiento adecuada a la población objetivo de los estudios evaluados. |

IA: Inteligencia artificial

REFERENCIAS BIBLIOGRÁFICAS

- Zhang P, Kamel Boulos MN. Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. *Future Internet*. 2023;15(9):286. doi: 10.3390/fi15090286.
- Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digit Med*. 2023;6(1):113. doi: 10.1038/s41746-023-00858-z.
- Al Barajraji M, Niset A, Englebert A, El Hadwe S, Barrit S. Beyond peer review: rethinking scientific publishing with artificial intelligence. *Intensive Care Med*. 2024;50(10):1715-6. doi: 10.1007/s00134-024-07559-5.
- International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals [Internet]. ICMJE; 2023 [citado el 8 de octubre de 2024]. Disponible en: <https://www.icmje.org/recommendations/browse/>
- World Association of Medical Editors. Chatbots, Generative AI, and Scholarly Manuscripts [Internet]. WAME; 2023 [citado el 8 de octubre de 2024]. Disponible en: <https://wame.org/page3.php?id=106>
- Committee on Publication Ethics. Authorship and AI tools [Internet]. COPE; 2023 [citado el 8 de octubre de 2024]. Disponible en: <https://publicationethics.org/cope-position-statements/ai-author>
- Checco A, Bracciale L, Loreti P, Pinfield S, Bianchi G. AI-assisted peer review. *Humanit Soc Sci Commun*. 2021;8(1):1-11. doi: 10.1057/s41599-020-00703-8.
- Cheng K, Sun Z, Liu X, Wu H, Li C. Generative artificial intelligence is infiltrating peer review process. *Crit Care*. 2024;28(1):149. doi: 10.1186/s13054-024-04933-z.
- Biswas S, Dobaría D, Cohen HL. ChatGPT and the Future of Journal Reviews: A Feasibility Study. *Yale J Biol Med*. 2023;96(3):415-20. doi: 10.59249/SKDH9286.
- Liang W, Izzo Z, Zhang Y, Lepp H, Cao H, Zhao X, *et al*. Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews [Internet]. arXiv.org; 2024 [citado el 8 de octubre de 2024]. Disponible en: <https://arxiv.org/abs/2403.07183v2>
- Saad A, Jenko N, Ariyaratne S, Birch N, Iyengar KP, Davies AM, *et al*. Exploring the potential of ChatGPT in the peer review process: An observational study. *Diabetes Metab Syndr*. 2024;18(2):102946. doi: 10.1016/j.dsx.2024.102946.
- Verharen JPH. ChatGPT identifies gender disparities in scientific peer review. *Elife*. 2023;12:RP90230. doi: 10.7554/eLife.90230.
- Munafò M. A Policy on the Use of Artificial Intelligence and Large Language Models in Peer Review. *Nicotine Tob Res*. 2024;26(5):519. doi: 10.1093/ntr/ntad242.
- Kaebnick GE, Magnus DC, Kao A, Hosseini M, Resnik D, Dubljević V, *et al*. Editors' Statement on the Responsible Use of Generative AI Technologies in Scholarly Journal Publishing. *AJOB Neurosci*. 2023;14(4):337-40. doi: 10.1080/21507740.2023.2257181.
- Ornes S. Peering inside the black box of AI. *Proc Natl Acad Sci U S A*. 2023;120(22):e2307432120. doi: 10.1073/pnas.2307432120.
- Hatem R, Simmons B, Thornton JE. A Call to Address AI "Hallucinations" and How Healthcare Professionals Can Mitigate Their Risks. *Cureus*. 2023;15(9):e44720. doi: 10.7759/cureus.44720.
- Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. *Commun Med*. 2021;1(1):1-3. doi: 10.1038/s43856-021-00028-w.
- Silva T. Necropolítica algorítmica. *Sex, Salud Soc (Rio J)*. 2023;(39):e22304. doi: 10.1590/1984-6487.sess.2023.39.e22304.a.es.
- Batista D. Tarcízio Silva: "O racismo algorítmico é uma espécie de atualização do racismo estrutural" [Internet]. Rio de Janeiro: CEE Fiocruz; 2023 [citado el 8 de octubre de 2024]. Disponible en: <https://cee.fiocruz.br/?q=Tarcizio-Silva-O-racismo-algoritmico-e-uma-especie-de-atualizacao-do-racismo-estrutural>
- Suleiman A, von Wedel D, Munoz-Acuna R, Redaelli S, Santarisi A, Seibold E-L, *et al*. Assessing ChatGPT's ability to emulate human reviewers in scientific research: A descriptive and qualitative approach. *Comput Methods Programs Biomed*. 2024;254:108313. doi: 10.1016/j.cmpb.2024.108313.
- Dutrénit G, Aguirre-Bastos C, Puchet M, Salazar M. Chapter 7: Latin America. En: UNESCO science report: the race against time for smarter development [Internet]. Ginebra: UNESCO; 2021 [citado el 8 de octubre de 2024]. p. 201-33. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000377462.locale=en>
- Silva CNN da. A confiança na ciência: critérios e condições para seu fortalecimento. *Rev. Nova Paid.*. 2020;2(2):1-3. doi: 10.36732/riep.v2i2.70.
- Ministerio de Salud. Prioridades Nacionales en Salud 2024-2030. Resolución Ministerial N.º 184-2024-MINSA [Internet]. Lima: MINSA; 2024 [citado el 8 de octubre de 2024]. Disponible en: <https://www.gob.pe/institucion/minsa/normas-legales/5364816-184-2024-minsa>
- Espinoza Vidaurre SM, Velásquez Rodríguez NC, Gambetta Quelopana RL, Martínez Valdivia AN, Leo Rossi EA. Influencia de la Inteligencia Artificial en la Eficiencia del Rendimiento Académico: Un Análisis de Determinantes. *Revista Ibérica de Sistemas e Tecnologías de Informação*. 2024;(70):399-418.
- Anales de la Facultad de Medicina. Instrucciones para los autores [Internet]. 2024 [citado el 8 de octubre de 2024]. Disponible en: <https://revistasinvestigacion.unmsm.edu.pe/index.php/anales/Instrucciones>
- Revista de Neuro-Psiquiatría. Políticas éticas de publicación [Internet]. 2023 [citado el 8 de octubre de 2024]. Disponible en: https://revistas.upch.edu.pe/index.php/RNP/politicas_eticas