

Andes Pediatría y publicaciones científicas en la era de la Inteligencia Artificial

Andes Pediatría and scientific publications in the era of Artificial Intelligence

Francisco Cano Sch.^a, Luisa Schonhaut B.^a, Paul R. Harris^a, Teresa Millán K.^a,
Patricia Mena N.^a, Macarena Lizama C.^a, Carolina Heresi V.^a, Pablo Cruces R.^a,
Ana Zepeda O.^a, Lillian Bolte M.^a, Macarena Gajardo Z.^a, Gerardo Weisstaub^a

^aComité Editorial, Andes Pediatría. Chile.

La Inteligencia Artificial (IA) es un término que se utiliza para definir un sistema informático creado y entrenado por humanos para desarrollar las capacidades de almacenar información o datos, interpretarlos, aprender de ellos, y aplicar el aprendizaje adquirido para llevar a cabo diversas acciones. Cuando estas acciones están orientadas a resolver procesos o problemas específicos, hablamos de IA específica, como, por ejemplo, un *software* de reconocimiento facial, y cuando el sistema informático es capaz de resolver temas en base a grandes bases de datos que abarcan cualquier materia, generando un razonamiento similar al humano, hablamos de IA general.

La IA se basa en complejos modelos lingüísticos, en inglés *Large Language Models (LLM)*, los cuales son algoritmos matemáticos que se utilizan para comprender y procesar el lenguaje natural. Se generan y entrenan en grandes conjuntos de datos (*big data*) y se utilizan para diversas tareas, como la generación y clasificación de textos, análisis de temas específicos, respuesta a preguntas, traducción de textos, entre otros. La generación de respuestas y textos coherentes sigue a la presentación de una pregunta (*prompts*), que corresponde a una secuencia de texto o imágenes que IA utilizará como instrucción para analizar, interpretar y generar una respuesta coherente y relevante. Dado

que el algoritmo desarrolla la respuesta en base a información producida por un grupo de seres humanos, el resultado tendrá importantes similitudes con el lenguaje humano, y como usa un espectro de información y análisis muy superior al que puede usar una sola persona, es previsible que el resultado supere ampliamente lo que usuarios estudiantes o investigadores puedan producir.

La IA, en permanente evolución y perfeccionamiento, será capaz de un auto-entrenamiento y proposición de ideas generadas en un contexto artificial, muy lejos de las actuales prestaciones como son la generación de textos, reconocimiento de voz, reconocimientos faciales, generación de gráficos y cálculos matemáticos avanzados, entre otras. La aplicación más utilizada de la IA es aquella que se produce mediante la interacción directa entre un ser humano y el *chatbot*, denominación genérica que recibe el programa de IA al interactuar con el ser humano, el equivalente a un robot pero sin presencia física. Entre los LLM o GTP (*Generative Pre-trained Transformer*) más conocidos están ChatGPT (chat.openai.com/auth/login), GPT (openai.com/gpt-4), Open AI (openai.com), Bard (bard.google.com/?hl=es), Claude (claude.ai/login), entre otros.

Dada la similitud en el lenguaje usado por el *chatbot* y el texto escrito por un ser humano, se hace ne-

Correspondencia:
Francisco Cano Sch.
fcanosch@gmail.com

cesario diferenciar entre ambos. Para esto se ha usado tradicionalmente el Test de Turing, nombrado así en honor al matemático británico considerado por muchos como el creador de la IA. En este test, un experto humano debía ser capaz de distinguir 2 conversaciones, una generada por un humano, y la otra por una máquina. Si el evaluador no era capaz de diferenciarlas, la máquina o el algoritmo había superado la prueba y podía denominarse IA.

La pregunta que debe responder el mundo académico es hasta donde estos algoritmos son capaces hoy, y serán capaces en el futuro, de generar textos científicos que puedan aprobar el Test de Turing, respondiendo cuestiones técnicas superando la capacidad del razonamiento humano, reemplazando de esta forma al investigador, al científico, al médico, u otros profesionales.

Un buen ejemplo de la capacidad de estos modelos matemáticos es el experimento realizado por investigadores del Massachusetts General Hospital el año 2022¹, para evaluar la potencialidad de *ChatGPT* frente al examen USMLE (*United States Medical Licensing Exam*), el cual consta de tres pruebas estandarizadas que los estudiantes de medicina de Estados Unidos deben aprobar para obtener la licencia médica. Los investigadores formularon 350 preguntas abiertas y de opción múltiple, comprobando que el chatbot alcanzaba el umbral de aprobación del 60% requerido para aprobar el examen. Este experimento ha sido replicado en otras instancias médicas académicas con resultados variables, como el estudio multicéntrico de la Universidad de Yale, *New Haven, USA*, y la Universidad Nacional de Dublin, Irlanda, quienes aplicaron el *United States Medical Licensing Examination Step 1 and Step 2*, en un formato de 4 secciones de diverso grado de dificultad. *ChatGPT* alcanzó un puntaje de 44% (44/100), 42% (42/100), 64,4% (56/87), y 57,8% (59/102), en cada una de las secciones². El modelo demostró una disminución significativa del rendimiento a medida que aumentaba la dificultad de las preguntas, lo cual ha sido comprobado igualmente al aplicar exámenes médicos en subespecialidades como los test nefrológicos KSAP y NephSAP³.

Otro ejemplo, es el experimento de Gao y cols.⁴, quienes recopilaron 50 resúmenes de cinco revistas médicas de alto factor de impacto, y solicitaron a *ChatGPT* que generara resúmenes de investigación basándose en los títulos de la investigación y revistas en la que fueron publicados, entregándolos luego a científicos, detectores de plagio y detectores de uso de IA. En la investigación reportaron que los resúmenes elaborados a través de IA tenían menos puntajes en plagio que los escritos por los investigadores. Por otro lado, tanto los humanos como los detectores de IA eran capaces de identificar una parte de los resúmenes generados por

ChatGPT. Sin embargo, ninguno de ellos era un discriminador perfecto.

En el campo de la medicina clínica, los *chatbot* cambiarán la forma en que ésta se ha conducido a lo largo de la historia. La IA se perfila como una herramienta de alto impacto en la atención del paciente, con un ilimitado acceso a la información disponible en las bases de datos, por la capacidad de generar correlaciones entre las innumerables experiencias publicadas, aplicando herramientas de selección según solicite el médico tratante, y muy probablemente en el futuro próximo, generando sugerencias y conductas de acuerdo a lo analizado, constituyendo así una forma de apoyo a los profesionales los cuales deberán manejar una herramienta que, sobrepasando ampliamente las capacidades humanas, requiere un aprendizaje y un manejo íntegro, técnico y ético. Su uso fuera de un contexto profesional responsable plantea profunda inquietud en el mundo de la medicina, de la ética y de la filosofía⁵. Por otro lado, los *chatbot* también representan una forma de asesoramiento médico al paciente que sin duda será ampliamente utilizado, lo cual plantea importantes consideraciones por parte de los profesionales a cargo, modificando una interacción tradicionalmente vertical y que sin duda cambiará la relación médico-paciente.

En el campo de la publicación científica, la IA incluye herramientas de ayuda a la redacción, revisión gramatical y de lenguaje, apoyo a las referencias, análisis estadístico y ajuste a las normas de publicación. Los editores de revistas biomédicas han manifestado su preocupación por la capacidad de la IA de generar textos humanizados en base a la información almacenada en millones de artículos disponibles en bases de datos como Medline o Scopus, y accesibles a través de potentes buscadores como PubMed. La generación de texto inteligente ha dado el nombre comúnmente usado de LLM a estos algoritmos, actualmente accesibles vía inscripción pagada, o en versiones básicas libres, como *ChatGPT*. El uso de estos *chatbots* permite generar textos complejos que un investigador puede utilizar para complementar su estudio y apoyar la elaboración del manuscrito final, pero que puede igualmente ser usado para generar un manuscrito prácticamente sin intervención humana.

Al revisar las normas de autoría de la ICMJE (International Committee of Medical Journal Editors)⁶ podemos considerar como autor de un manuscrito a quien apoya la redacción del mismo. Desde ese punto de vista, ha sido discutido hasta qué punto la IA podría ser incluida en la lista de coautores, lo que ha sido descartado debido a que estos algoritmos no pueden dar cuenta responsablemente de lo redactado, pero es fundamental declararlo al momento de publicar el manuscrito⁷. En ese sentido, se solicita que cada sección escrita con apoyo de IA sea críticamente revisada por el

equipo de investigación para verificar su precisión, sesgo, relevancia y razonamiento, considerando que esta es una herramienta nueva, que no reemplaza el rol del investigador y el pensamiento científico, y que requiere ser declarada al momento de la publicación⁸.

Desde etapas previas a la educación universitaria, el abuso de estas herramientas puede distorsionar el aprendizaje de estudiantes y futuros investigadores limitando su pensamiento crítico, iniciativa y creatividad, impactando en el desarrollo de los Recursos Mentales, término usado por Capponi en su libro *Felicidad Sólida*, y que representa la capacidad de análisis, de razonamiento lógico, y de creatividad del cerebro humano. Estos Recursos Mentales son la base del desarrollo y evolución humana⁹. Sin embargo, si *ChatGPT* se utiliza adecuadamente puede ser una herramienta que ayude a desarrollar y potenciar los Recursos Mentales a un nivel nunca antes vistos. Comprender el impacto de las herramientas que utilizan IA en el proceso de enseñanza/aprendizaje es un desafío para el sistema educativo, no sólo a nivel del desarrollo de competencias de los estudiantes en cada etapa de la vida, sino también en el aporte a la evolución de la sociedad en su conjunto.

Se ha citado como analogía el uso de las calculadoras en el aprendizaje de las matemáticas. Cuando las calculadoras se pusieron a disposición de los estudiantes desde sus primeras etapas escolares, se temió fuese una limitante al aprendizaje y progreso de las habilidades en ese campo. Sin embargo, la incorporación de esta herramienta, su aplicación masiva, dio una nueva base al aprendizaje, la disciplina se adaptó y las calculadoras llevaron la enseñanza y la aplicación de las matemáticas a un nuevo nivel cognitivo. Eso lo vemos hoy en nuestros alumnos de pre y postgrado. Esa misma lógica podemos aplicarla a los científicos que integren softwares como *ChatGPT* de forma que les permita desarrollar su pensamiento en un nivel más avanzado que facilite el laborioso trabajo que representa la recolección, análisis de datos, generación de resultados y su interpretación, todos ellos pasos básicos de la investigación y publicación científica. Sin embargo, si el *software* no se utiliza de forma innovadora y responsable, los científicos derivarán hacia la mediocridad, porque su trabajo perderá la creatividad que sólo puede provenir de un pensamiento crítico original.

La delgada línea que separa el uso ético de un uso fraudulento debe ser motivo de preocupación y de la implementación de políticas editoriales que aseguren la veracidad de las investigaciones y de su producto final, el artículo científico, que en definitiva tendrá directo impacto en el paciente. Andes Pediátrica (AP), como todas las publicaciones científicas en el futuro, requiere generar una política de integración de este nuevo paradigma a la publicación científica, que ha

sido incorporada en distintas revistas de corriente principal^{7,10}. Sin duda alguna estamos entrando en una etapa hiperacelerada que, en especial los jóvenes, actual generación digital, deberán adaptarse para no quedar fuera de un contexto laboral, social y profesional en un futuro cercano. Un primer paso será entender que muchos manuscritos incorporarán los *chatbot* como parte de su elaboración, para lo cual AP abrirá la opción de acceso a los artículos solicitando una declaración explícita de su uso, declarando aquellas secciones que han sido intervenidas por IA (<https://www.revistachilenadepediatria.cl/index.php/rchped/about/submissions#authorGuidelines>). La respuesta de AP a este nuevo paradigma será estimular el uso responsable y ético de la IA entre los investigadores en la generación de su investigación, en un proceso editorial que incluya la evaluación de esos artículos en una forma propositiva.

Una segunda etapa, para asegurar el estricto control científico y ético del material publicado, será establecer un control de los manuscritos mediante detectores de textos generados por IA, y en forma anecdótica, los editores de algún modo serán los que apliquen el test de Turing. Los revisores, parte fundamental del proceso editorial, deberán ser parte de estos tiempos de cambio, asegurando la anonimización del proceso editorial, y aplicando las potencialidades de la IA en su labor de revisión. La reciente jornada de editores y revisores sobre IA que AP desarrolló el mes de septiembre 2023, que incluyó bioinformáticos, especialistas en IA, y un miembro activo de la *National Library of Medicine* que mostró el enfoque a futuro basado en IA de este organismo, fue un punto de inflexión para entender que estamos entrando en una nueva era.

Probablemente no está lejos el momento en que los editores de las revistas científicas incorporen la evaluación de los manuscritos por la IA en calidad de revisor, algoritmos que evaluarán la originalidad del proyecto, la pertinencia de la pregunta de investigación, el planteamiento del objetivo, la exactitud de los datos y de la metodología usada, de los resultados, los alcances de la discusión y de las conclusiones, con mucha mayor eficiencia que un revisor humano, todo ello producto de un análisis basado en millones de documentos integrados por el *chatbot*.

Al momento de escribir esta editorial le preguntamos a *ChatGPT* sobre “¿cómo deben ser las políticas editoriales de revistas científicas en relación al uso de inteligencia artificial en la generación de manuscritos científicos?” y en el recuadro 1 reproducimos el texto generado como respuesta. Probablemente si le volvemos a preguntar en un tiempo, la respuesta sea diferente, o al menos con otros matices, pero da una buena cuenta de las políticas editoriales que implementaremos en AP.

Una reflexión final

El lenguaje es la base de la civilización humana. Mediante el lenguaje el Homo sapiens ha podido construir las historias en que se basa la sociedad humana. Heidegger, uno de los filósofos más influyentes del siglo XX, dijo que el lenguaje no es solo un medio de comunicación, sino que también es el fundamento de nuestra comprensión del mundo y de nuestra existencia. Sin embargo, agregó que el lenguaje puede ser problemático, ya que puede ocultar la verdad y llevarnos a una comprensión superficial de las cosas. Según el historiador israelí Yuval Noah Harari, autor de *Homo Sapiens*¹¹ y *Homo Deus*¹², es mediante historias comunes creadas por el lenguaje, que la sociedad ha podido alcanzar el nivel de desarrollo actual. Miles y millones de seres humanos creen en reglas de convivencia y conductas sociales, millones de humanos creen en dioses y siguen una religión con la fe en una existencia superior, millones de seres se organizan en torno a objetivos económi-

cos, se agrupan en naciones y banderas, y nada de ello tiene una base orgánica o biológica, todas estas historias son mitos compartidos, todos ellos han sido creados por el lenguaje. El lenguaje según Harari, es la llave maestra de la civilización humana, y su planteamiento es claro. El lenguaje ha sido entregado a máquinas infinitamente superiores a la capacidad del ser humano, por eso se denominan LLM, y estos algoritmos en una espiral de autoperfeccionamiento tendrán en muy poco tiempo el dominio de esta llave maestra. ¿Hasta dónde podrán estos algoritmos desarrollar sus capacidades y controlar las historias que dan forma a la sociedad? Esa, según Harari, es la gran pregunta. La especie *Homo Sapiens*, como la conocemos hoy, según este historiador y filósofo, está entrando en una nueva era.

Conflicto de interés

Los autores son miembros del Comité Editorial de Andes Pediátrica.

Le preguntamos a ChatGPT:¹³
<https://openai.com/chatgpt>



“¿cómo deben ser las políticas editoriales de revistas científicas en relación al uso de inteligencia artificial en la generación de manuscritos científicos?”

Esto fue lo que nos respondió:

“Las políticas editoriales de revistas científicas en relación al uso de inteligencia artificial (IA) en la generación de manuscritos científicos deben ser cuidadosamente diseñadas para abordar los desafíos y oportunidades que esta tecnología plantea. Aquí tienes algunas pautas para desarrollar políticas editoriales efectivas:

1. **Transparencia y divulgación de IA:** Exige que los autores revelen de manera explícita si han utilizado IA en la generación de su manuscrito. Esto debe incluir detalles sobre las herramientas, modelos y algoritmos utilizados.
2. **Evaluación y revisión por pares especializados:** Asegúrate de contar con revisores que estén familiarizados con el uso de IA en investigación científica. Los revisores deben evaluar no solo la calidad del trabajo, sino también la idoneidad y el rigor de los métodos de IA utilizados.
3. **Ética y responsabilidad:** Establece pautas claras sobre la ética en el uso de IA, incluyendo la recopilación y el procesamiento de datos. Los autores deben ser conscientes de la importancia de la ética en la investigación de IA y la toma de decisiones algorítmicas.
4. **Propiedad intelectual:** Aborda la cuestión de la propiedad intelectual en el caso de que una IA haya contribuido significativamente a la generación del manuscrito. ¿Quién es el autor o titular de los derechos de autor en este caso?
5. **Transparencia en los métodos de IA:** Exige que los autores proporcionen detalles sobre la configuración de la IA, los hiperparámetros, los conjuntos de datos utilizados y cualquier ajuste que hayan realizado. Esto es esencial para la reproducibilidad de los resultados.
6. **Verificación y validación:** Los autores deben llevar a cabo una verificación y validación rigurosa de los resultados generados por la IA para garantizar su precisión y fiabilidad. Esto es especialmente importante en campos críticos como la medicina.
7. **Acceso a código fuente y datos:** Fomenta la publicación del código fuente y los conjuntos de datos utilizados en la generación de los manuscritos. Esto facilitará la revisión y la reproducción de los resultados por parte de otros investigadores.
8. **Revisión ética y legal:** Considera establecer un proceso de revisión ética y legal para manuscritos generados por IA, especialmente si involucran datos sensibles o tienen implicaciones legales.
9. **Educación y concienciación:** Proporciona orientación a los autores sobre las mejores prácticas en el uso de IA en investigación científica y su responsabilidad ética.
10. **Actualización constante:** Dado que la tecnología de IA está en constante evolución, las políticas editoriales deben revisarse y actualizarse periódicamente para reflejar los avances y desafíos emergentes.

Al implementar estas recomendaciones en las políticas editoriales, las revistas científicas pueden abordar de manera efectiva la creciente influencia de la IA en la generación de manuscritos y garantizar la calidad, la integridad y la ética en la publicación de investigaciones impulsadas por IA.”

Referencias

1. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023; 2(2): e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.
2. Gilson A, Safranek CW, Huang T, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med Educ*. 2023;9:e45312. doi: 10.2196/45312.
3. Miao J, Thongprayoon C, Cheungpasitporn W. Assessing the Accuracy of ChatGPT on Core Questions in Glomerular Disease. *Kidney Int Rep*. 2023;8(8):1657-1659. doi: 10.1016/j.ekir.2023.05.014.
4. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digit Med*. 2023;6(1):75. doi: 10.1038/s41746-023-00819-6.
5. Garcia Valencia OA, Suppadungsuk S, Thongprayoon C, et al. Ethical Implications of Chatbot Utilization in Nephrology. *J Pers Med*. 2023; 13(9):1363. <https://doi.org/10.3390/jpm13091363>.
6. Internacional Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Defining the Role of Authors and Contributors. (Accedido 16.09.23), Disponible en: <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>
7. Abbasi K. Human, all too human - why artificial intelligence cannot "author" papers. *J R Soc Med*. 2023;116(5):159. doi: 10.1177/01410768231179306.
8. Hosseini M, Rasmussen LM, Resnik DB. Using AI to write scholarly publications. *Account Res*. 2023;1-9. doi: 10.1080/08989621.2023.2168535.
9. Capponi R. Felicidad Sólida. Primera edición, Caligrama 2019. ISBN 9788417120610; ISBN eBook 9788417533960.
10. Flanagan A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman "Authors" and Implications for the Integrity of Scientific Publication and Medical Knowledge. *JAMA*. 2023;329(8):637-9. doi:10.1001/jama.2023.1344.
11. *ChatGPT* (versión 16 septiembre 2023) [Large language model].
12. Harari Y N. *Sapiens: a Brief History of Humankind*. New York :Harper, 2015. ISBN 9780062316097.
13. Harari Y N. *Homo Deus*. Harvill Secker, 2016. ISBN 9781910701881.