

## Implicaciones y políticas editoriales de la inteligencia artificial

[en] Implications and editorial policies of artificial intelligence

  Juan D. Machin-Mastromatteo<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Chihuahua (México)

Recibido: 2023/09/19

Aceptado para su publicación: 2023/11/05

Publicado: 2023/12/15

### RESUMEN

En esta entrega de *Escuela de editores* presento algunas de las problemáticas alrededor del uso de la inteligencia artificial (IA) en la investigación para enmarcar la síntesis y análisis de diez políticas editoriales sobre el uso de IA: arXiv, Elsevier, Emerald Publishing, International Committee of Medical Journal Editors, Oxford University Press, Sage, Springer Nature, Taylor & Francis Group, Wiley y World Association of Medical Editors. Abordo cómo los modelos de lenguaje de gran tamaño, como *ChatGPT*, han afectado la integridad y originalidad de los productos científicos, examinando problemas como el entrenamiento de IA, su uso en la autoría de documentos científicos, la veracidad del contenido generado por IA y el impacto que están teniendo estas tecnologías en la investigación y publicación científica. Las mencionadas políticas editoriales actuales sobre el uso de IA enfatizan la necesidad de actuar con transparencia sobre su uso, además de subrayar la responsabilidad de los autores y las limitaciones éticas y prácticas de usar IA en la investigación. Concluyo con una reflexión acerca de buscar un equilibrio entre las ventajas y limitaciones de la IA para poder emplearla en la investigación y publicación científica sin faltar a valores de ética e integridad científicas.

### PALABRAS CLAVE

Inteligencia artificial, modelos de lenguaje de gran tamaño, chatbots, investigación científica, escritura científica, publicación científica, políticas editoriales, directrices editoriales, ética en la investigación, integridad en la investigación.

### Cómo citar (APA 7<sup>a</sup> Edición):

Machin-Mastromatteo, J. D. (2023). Implicaciones y políticas editoriales de la inteligencia artificial. *Revista Estudios de la Información*, 1(2), 123-133. <https://doi.org/10.54167/rei.v1i2.1448>

### ABSTRACT

In this entry of the *School of editors* section, I present some of the issues surrounding the use of artificial intelligence (AI) in research to frame the synthesis and analysis of ten editorial policies on the use of AI: arXiv, Elsevier, Emerald Publishing, International Committee of Medical Journal Editors, Oxford University Press, Sage, Springer Nature, Taylor & Francis Group, Wiley and World Association of Medical Editors. I address how large language models, such as *ChatGPT*, have affected the integrity and originality of scientific products, examining issues such as AI training, its use in scientific authoring, the accuracy of AI-generated content, and the impact these technologies are having on scientific research and publishing. Current editorial policies on the use of AI emphasize the need to act with transparency if they are used, in addition to stressing authors' responsibilities and the ethical and practical limitations of using AI in research. I conclude with a reflection about seeking a balance between the advantages and limitations of AI in order to be able to use it in scientific research and publication without compromising ethics and integrity values in science.

### KEYWORDS

Artificial intelligence, large language models, chatbots, scientific research, scientific writing, scientific publishing, editorial policies, editorial guidelines, research ethics, research integrity.

## Introducción

La inteligencia artificial (IA) fue uno de los temas más populares de 2023 y ha sido fascinante ver cómo ha permeado en cualquier actividad. En la educación se ha cuestionado si hará los trabajos de la escuela y de manera similar también algunos investigadores o candidatos a investigadores podrían pensar que ahora la IA les va a escribir sus artículos, pues de hecho no es así y hay múltiples consideraciones que hacer al respecto. En esta entrega de *Escuela de Editores* presento aquellas que están relacionadas con las políticas editoriales que han surgido como respuesta a la masificación y altos niveles de adopción de esta tecnología.

Es importante aclarar que, por practicidad, a la tecnología a la cual nos referiremos como IA, que es aquella que tuvo un amplio auge y sobre la cual se habló casi hasta el hartazgo a lo largo de 2023, es en realidad un tipo muy específico de IA: los llamados modelos de lenguaje de gran tamaño. Estos son modelos probabilísticos que se basan en procesos de aprendizaje profundo, aprendizaje automático, redes neuronales, procesamiento de lenguaje natural y modelado generativo. Como toda *IA estrecha*, estas se especializan en realizar una única tarea específica: generar textos con cierta coherencia como respuestas a preguntas o instrucciones (*prompts*) realizadas por sus usuarios; al contrario de lo que sería una *IA general*, que se especializaría en múltiples actividades y la cual al día de hoy no existe. Adicionalmente, existen varias soluciones de este tipo de IA, también conocidas como *chatbots*, que compiten entre sí, donde el más famoso resultó ser *ChatGPT*, ya en su cuarta versión al momento de escribir este texto, pero también están disponibles sistemas homólogos como *Bard* y *Claude*.

## Problemáticas alrededor del uso de la inteligencia artificial en la investigación

Al margen de las amplias controversias del impacto de la IA en la educación, específicamente en cuanto a la integridad y originalidad de los trabajos que presenten los estudiantes, lo cual amerita amplias discusiones, pero que superan el alcance de este trabajo, a lo largo de 2023 fueron ocurriendo ciertos hechos que resaltaron la urgencia de realizar cambios en las políticas editoriales de actores relacionados con la producción y publicación científica. Antes de discutir las políticas editoriales en cuanto a IA, es relevante destacar los siguientes acontecimientos, que no conforman una lista exhaustiva, pero que permiten ilustrar ciertas problemáticas.

### ***Entrenamiento de las inteligencias artificiales y su base de conocimiento***

En términos sencillos, las IA han sido *entrenadas* con procesos manuales y automáticos que se basan en la revisión por parte de esta tecnología de grandes cantidades de documentos, para que pudiera *aprender a hablar*. Como tal, estas IA han sido caracterizadas utilizando las siguientes metáforas: (1) *loros estocásticos*, porque repiten como loro aquello que de alguna manera han aprendido, pero sin entenderlo o identificar correctamente de dónde proviene exactamente lo que dicen, además de que su discurso se desarrolla de manera probabilística, según ciertas palabras puedan conformar una secuencia dado un contexto o pregunta; y (2) *androides enjaulados en una biblioteca*, ya que se les ha encerrado con una serie de documentos para que los *lean* y así aprendan a *hablar* (Villegas-Ceballos, 2023).

Podemos referirnos a la documentación empleada en el entrenamiento de las IA como un *corpus* o como su *base de conocimiento*. La cual hasta cierto momento estuvo limitada obviamente por un volumen determinado, pero también bajo ciertos rangos de fechas. Esto ha venido cambiando desde que las IA empezaron a tener acceso a contenidos Web. Sin embargo, en ninguno de los casos un usuario común y posiblemente ni siquiera sus responsables podemos dimensionar, ni mucho menos caracterizar las fuentes que conforman tal base de conocimiento. A esto se le suma que, dentro del carácter probabilístico de sus respuestas, prácticamente aleatorio, no es posible distinguir cuáles fuentes ha empleado para construir una respuesta. Son auténticas *cajas negras*, ya que no podemos saber cuáles son aquellos documentos que se usaron para su entrenamiento, ni cuáles documentos está empleando en un momento determinado para darme una respuesta, dado que también sus respuestas a la misma pregunta podrían ser diferentes cada vez que se le presente un cuestionamiento o *prompt* específico. Sí,

se le puede pedir que cite fuentes, pero ¿cómo podríamos estar seguros de que lo ha hecho de manera exhaustiva (como requiere hacerse en el discurso científico)? Además, se ha determinado que inventa tanto citas, como fuentes.

### ***Las inteligencias artificiales como autores de documentos científicos***

Un *preprint* es una versión preliminar de un documento científico (usualmente un artículo, aunque también podría tratarse de un libro, capítulo o protocolo de laboratorio), que se deposita en un *servidor de preprints* o un repositorio para compartir sus resultados lo antes posible con la comunidad científica. Esto implica que un *preprint* no ha sido revisado por pares y puede que haya sido enviado también a una revista para su posible publicación. Kung et al. (2022) publicaron en *MedRxiv*, el mayor servidor de preprints dedicado a las áreas de la salud, uno de posiblemente varios *preprints* que enlistaba una IA, específicamente a *ChatGPT*, como coautor del documento. Posteriormente, Kung et al. (2023) publicaron la versión final de su documento en *PLOS Digital Health*, omitiendo la coautoría del sistema mencionado.

### ***Rastros del uso de una inteligencia artificial***

En la revista *Toxins* de la editorial *Multidisciplinary Digital Publishing Institute* (MDPI), el artículo de Tsai et al. (2023) presentó el peculiar subtítulo “*Regenerate Response*” dentro de la discusión. Aunque el documento fue *corregido* y este texto fue eliminado, todavía puede consultarse la primera versión con tal subtítulo, que presuntamente fue copiado junto a una respuesta de *ChatGPT*, dado que debajo de cada respuesta de este sistema aparecía un botón con ese texto, para pedir al sistema que generara otra respuesta a la pregunta que se le había hecho anteriormente. Al momento de escribir este texto, se sustituyó este botón por un ícono de una flecha circular, que usualmente ilustra la opción de *actualizar* en distintos sistemas.

Este tipo de problemática ha señalado una posible línea de investigación: analizar los rastros del uso de una IA en publicaciones científicas. Un ejemplo es el trabajo del que da cuenta el preprint de Tang y Eaton (2023), que, aunque analiza una muestra muy pequeña, propone una idea interesante: analizar las características de artículos que presenten frases como “*Como un modelo de lenguaje*” (“*As an AI language model*”), típicas de estas IA como una forma de iniciar sus respuestas, generalmente cuando se le está pidiendo una opinión. Los autores se refieren a estos fragmentos como *huellas de contenido generado por IA*.

### ***Alucinaciones de la inteligencia artificial***

La veracidad del contenido generado por IA es cuestionada, recordemos que al definir esta tecnología usé la palabra coherencia, no veracidad. Ciertamente, el contenido que genera tiene cierto sentido a nivel de presentación e incluso lingüístico, pero puede ser incorrecto o impreciso. Según Spinak (2023), un modelo de lenguaje “percibe patrones que son inexistentes o imperceptibles para los observadores humanos, creando resultados inesperados o incorrectos” (párr. 1), a lo cual se le conoce como alucinación. Adicionalmente, Sage (2023) ilustra que tales alucinaciones:

generan contenido falso, especialmente cuando se emplean fuera de sus dominios o al lidiar con temas complejos o ambiguos. Pueden generar contenido que es plausible lingüísticamente pero no científicamente, pueden expresar hechos erróneos, han demostrado generar citas que no existen. Algunos modelos son solamente entrenados con contenidos publicados antes de una fecha particular y por lo tanto presentan un panorama incompleto (párr. 3).

Es por esto que usar IA para la investigación científica se trata de algo muy delicado y con lo que se debe tener mucho cuidado, dado que en la investigación es vital el poder confiar en la validez y precisión de la información que contiene un documento científico, así como en la correcta atribución de la autoría, la cual a su vez dota de solidez a la confiabilidad de la información. Incluso pareciera que cuando estos modelos de lenguaje no saben algo, pues inventan. Podemos sumar a las alucinaciones lo

que ocurre cuando pedimos a las IA que incorporen citas y referencias en las respuestas que nos vayan a ofrecer, donde puede ocurrir que efectivamente arroje una cita genuina, pero también puede atribuir una idea a un documento que no la presenta (muy problemático si se trata de una cita textual) o por el contrario puede que cite un trabajo de un experto en un tema, pero que tal trabajo no exista; problema que comento en la siguiente sección.

### ***El impacto de la inteligencia artificial***

Orduña Malea (2023) denunció en Twitter (X) que en varios *servidores de preprints* se estaban publicando documentos que citaban publicaciones donde él era coautor, pero el problema es que tales documentos no existen, nunca los ha publicado. Esto se vuelve algo más grave cuando sistemas como *Google Académico* y *ResearchGate*, que indizan documentos y sus redes de citas, registran estos preprints, ya que agregan estos documentos y las referencias que citan, sin distinguir los documentos que nunca han existido; al parecer no hacen tal comprobación. Orduña Malea (2023) se refiere a esto con cierta ironía como el *impacto de ChatGPT*, ya que es capaz de identificar correctamente a autoridades en un tema de investigación específico, pero ha llegado a recomendar documentos que pueden o no existir. Ciertamente, esto ha generado un nuevo problema bibliométrico en este tipo de sistemas, ya que les llevaría a agregar de manera automática a los perfiles de autores algunos documentos que no existen, pero que extrañamente han sido citados. Esto es especialmente problemático en *Google Académico*, que no tiene un mecanismo para sugerir correcciones y estas citas generadas por IA se suman a otros tipos de citas problemáticas, como las *citas spam* (Machin-Mastromatteo, 2022). Sin embargo, es justo decir que hay sistemas basados en IA, como *Consensus* y *Scite*, cuyas respuestas a nuestras preguntas consisten en una lista de resultados de documentos reales, resaltando fragmentos que pueden contener las respuestas que buscamos, además de marcar si estos documentos son revisiones sistemáticas o han sido muy citados.

### **Políticas editoriales sobre el uso de la inteligencia artificial**

Dadas estas problemáticas y el potencial disruptivo de la IA en la producción y publicación científica, no tardaron en aparecer a lo largo de 2023 ciertas modificaciones claves a las políticas de distintos actores relacionados con estas actividades, o incluso, que hayan surgido nuevas políticas al respecto de la IA. Estos cambios como respuesta a la extensión de su uso y para intentar prevenir malas prácticas. A continuación, presento un resumen y análisis de diez políticas editoriales, específicamente en cuanto a elementos relacionados con el uso de la IA en la investigación y publicación científica. Las políticas analizadas fueron, en orden alfabético, las de arXiv (2023), Elsevier (2023a, 2023b, 2023c), Emerald Publishing (2023), International Committee of Medical Journal Editors (ICMJE, 2023), Oxford University Press (OUP, 2023), Sage (2023), Springer Nature (2023), Taylor & Francis Group (T&FG, 2023), Wiley (2023) y de la World Association of Medical Editors (WAME, 2023).

### ***Reportar el uso de inteligencias artificiales generativas***

Todas las políticas analizadas señalan que los autores que empleen herramientas de IA para realizar su documento deben reportar su uso, así como señalar claramente cuál fue el contenido generado, el propósito de haber desarrollado el contenido de tal manera e indicar cuál herramienta se usó (arXiv, 2023; Elsevier, 2023a, 2023c; Emerald Publishing, 2023; ICMJE, 2023; OUP, 2023; Sage, 2023; Springer Nature, 2023; T&FG, 2023; WAME, 2023; Wiley, 2023). WAME (2023) es la más específica al mencionar que tal reporte debe incluir los *prompts* utilizados para generar contenido, la fecha y hora en que se realizó el intercambio con la IA, la herramienta empleada y su versión.

La mayoría de las políticas indican que el reporte del uso de IA debe incluirse en la sección de metodología de los artículos y alternativamente en la sección de reconocimientos o agradecimientos (*acknowledgements*), pero hay variantes: ICMJE (2023) requiere que se incluya en el texto del artículo y en su correspondiente carta de presentación (*cover letter*) enviada al editor, OUP (2023) pide que se incluya o bien en la metodología o en los *acknowledgements*, así como en la carta de presentación y

Elsevier (2023a) señala que debe incluirse antes de las referencias, en una sección titulada *Declaration of Generative AI and AI-assisted technologies in the writing process*.

En cuanto al propósito de emplear estas tecnologías, se hace complicado justificarlo de manera razonable, lo cual quizás estaría muy claro y sería entendible en los casos en que el objeto de estudio, o al menos parte de este, fuera la misma IA. De lo contrario, no tiene mucha razón de ser, ya que una posible objeción de cualquier revisor par podría consistir en cuestionar por qué parte del contenido no fue generado por humanos. WAME (2023) extiende el reportar el uso de IA al proceso de revisión de un documento, especificando que los editores y revisores pares deben declarar el uso de IA a los autores y también deben hacerlo entre editores y revisores, si es que se han utilizado para generar las revisiones e incluso la correspondencia que se intercambia entre todos estos actores a lo largo del proceso de publicación. Springer Nature (2023), aunque pide a los revisores pares que no utilicen IA para evaluar manuscritos, también dice que, si lo hacen, les solicita a los revisores que declaren su uso en el texto de sus evaluaciones.

### **La importancia de la autoría**

Las políticas reafirman la responsabilidad de los autores sobre la exactitud, integridad y originalidad de sus documentos (arXiv, 2023; Elsevier, 2023a, 2023c; Emerald Publishing, 2023; ICMJE, 2023; Springer Nature, 2023; T&FG, 2023; WAME, 2023; Wiley, 2023) y que como herramientas tecnológicas, las IA no pueden ser incluidas como autores (arXiv, 2023; Elsevier, 2023a, 2023c; ICMJE, 2023; OUP, 2023; Sage, 2023; Springer Nature, 2023; T&FG, 2023; WAME, 2023; Wiley, 2023), ya que no pueden asumir responsabilidad o dar consentimiento. De esto último se derivan dos palabras clave: responsabilidad y consentimiento. El *Committee on Publication Ethics* (COPE, 2023b), organización sin fines de lucro que se ha convertido en referente en cuanto a temas de ética e integridad de la investigación y publicación científica, sintetiza el por qué una IA no califica como autor de la siguiente manera:

Las herramientas de IA no pueden cumplir los requisitos de autoría, ya que no pueden asumir la responsabilidad del trabajo presentado. Como entidades no jurídicas, no pueden afirmar la presencia o ausencia de conflictos de intereses ni gestionar los derechos de autor y los acuerdos de licencia<sup>1</sup> (párr. 2).

Lo más sencillo es hablar primero del consentimiento, ya que para poder otorgarlo se requiere que un autor sea una persona como un ser consciente y que pueda aceptar que cierta versión de un documento sea por ejemplo enviada a una revista para su consideración a publicarse. Dado que una tecnología, por muy *inteligente* que se llame, hasta el momento no se considera un ser con conciencia y por lo tanto no puede consentir que se publique un trabajo donde tuvo algo que ver en su elaboración. Habrá sin embargo algún momento en que se cree una IA que desarrolle una conciencia propia y se darán interesantes debates filosóficos al respecto de qué puede o no decidir por sí misma.

El tema de la responsabilidad y en cierta medida el del consentimiento también nos remonta a quizás una de las directrices de autoría de investigación más antiguas, conocida popularmente como el *Protocolo Vancouver* (ICMJE, 2023<sup>2</sup>); el cual no hay que confundir con las normas de estilo de citación Vancouver. El *Protocolo Vancouver*, conocido originalmente como los *Uniform Requirements for Manuscripts Submitted to Biomedical Journals* y publicado por primera vez en 1978 establece, entre otras cosas, quién puede ser considerado como autor y sus responsabilidades.

El *Protocolo Vancouver* especifica que todo autor realizó *contribuciones sustanciales* para la concepción o diseño del trabajo, o para la recolección, análisis o interpretación de datos (al menos en una de estas etapas), además que trabajó en el primer borrador del documento o hizo una revisión crítica de este sugiriendo mejoras, aprobó la versión del trabajo que se envió a la revista y está de acuerdo en asumir

<sup>1</sup> Acuerdos a los que en español nos referimos comúnmente como los *acuerdos de publicación* que se celebran entre autores y revistas o entre autores y editoriales.

<sup>2</sup> Versión más reciente de mayo de 2023.

ciertas responsabilidades sobre la exactitud, integridad y originalidad de los documentos que publica a su nombre. Específicamente, podemos sintetizar las responsabilidades de los autores en tres tipos: (1) responsabilidad legal: por ejemplo, si al autor le toca responder por problemas de integridad y legales que pudiera ocasionar su documento publicado; (2) responsabilidad moral: en caso que su publicación le haga merecedor de méritos, reconocimientos o penalizaciones; y (3) responsabilidad económica: en cuanto a comprometerse a pagar los costos de publicación y si gracias a su publicación recibe premios, estímulos o subvenciones (ICMJE, 2023).

La mayoría de las directrices relacionadas con la ética y la integridad de la investigación y publicación científica no son infalibles, es decir, por sí solas no previenen que ocurran vicios o *patologías de la autoría* si las personas no las adoptan de una manera comprometida. El *Protocolo* permite explicar por qué no es buena idea incluir como autor a cualquier persona que no trabajó en la investigación o en el artículo, a pesar de las presiones políticas o personales que pueda ejercer sobre los demás autores. A pesar de esto ocurren vicios como la autoría fantasma (es decir, quien escribió el artículo no figura en la lista de autores), la autoría política (se incluyó a algún autor porque detenta alguna posición de poder, la cual ejerce para que otros le incluyan como autor), la autoría piadosa (por *apoyar* a alguien) o por devolver el favor de que otro nos haya incluido en sus publicaciones. Estas desviaciones patológicas deben evitarse, ya que, si algo bueno resulta de la publicación del artículo, todos los autores serían recompensados y en caso contrario, si la publicación genera controversias, todos deben asumir sus responsabilidades. Alguien que no haya hecho una contribución sustancial a la investigación o al texto que recoge sus resultados podría ser premiado o castigado injustamente por las características o consecuencias que ocasione el producto final.

### ***La inteligencia artificial como corrector de estilo, no como analista***

Pocas políticas de las analizadas establecen que las IA pueden usarse para editar y mejorar la redacción de un trabajo (ortografía y gramática), pero no para interpretar datos o extraer conclusiones (Elsevier, 2023a, 2023b, 2023c; Emerald Publishing, 2023; Sage, 2023; Wiley, 2023). Es decir, las políticas se oponen a que las IA se empleen para el análisis, limitándose a recomendar su uso exclusivamente para corregir cuestiones de forma, como la redacción, la ortografía, la gramática y el estilo. Emerald Publishing (2023) justifica este uso de la siguiente manera:

sería permisible ya que esto refleja las herramientas estándar que ya se emplean para mejorar la ortografía y gramática, y usa material existente creado por el autor, en lugar de generar contenido completamente nuevo, mientras que el autor continúa siendo responsable por el trabajo original (sección *Generative AI uSage key principles*, párr. 4).

Emerald Publishing (2023) parece tener la política más exigente, al no permitir generar ningún contenido. Específicamente indica que esta prohibición se relaciona con la generación de contenidos en cualquier parte de un artículo, incluyendo resúmenes, revisiones de literatura, reportes estadísticos, de resultados o la creación de imágenes con IA.

### ***Verificación humana***

Las políticas requieren que los autores verifiquen cuidadosamente la exactitud, integridad, validez y pertinencia del texto y también de las citas (para prevenir aquello de las citas y referencias inventadas por la IA), corrigiendo los errores o inconsistencias en cualquier contenido generado por IA (Elsevier, 2023c; Emerald, 2023; ICMJE, 2023; Sage, 2023; T&FG, 2023; WAME, 2023; Wiley, 2023); dado que los resultados generados por IA pueden ser incorrectos, incompletos o sesgados (arXiv, 2023; Elsevier, 2023a, 2023c; Sage, 2023). Esto último nos remite a las limitaciones de estas IA como indicamos anteriormente, específicamente en cuanto a las características de las fuentes (base de conocimiento) que se usaron para su entrenamiento y las que esté empleando en un momento determinado para dar una respuesta, que podría ser tendenciosa o incompleta, especialmente si no sabemos cuáles fuentes está utilizando, lo cual es característico del uso de las versiones gratuitas de herramientas como *ChatGPT*.

Aquí vale la pena especificar que las funciones sin costo de estas herramientas nos permiten hacer preguntas y obtener respuestas, mientras que una suscripción tanto a algunos de estos *chatbots*, como a sistemas que se han denominado *chatbots generales especializados* (por ejemplo, *Cody*, *ChatPDF* y *ChatDOC*) permiten tener un mayor control sobre la base de conocimiento, ya que se pueden hacer preguntas y obtener respuestas que se centren en uno o varios documentos que elija el usuario. Estos *chatbots generales especializados* a su vez utilizan los modelos de lenguaje como *ChatGPT*, por lo que, simplificando tecnicismos, podríamos verlos como interfaces que nos permiten delimitar el funcionamiento de los modelos.

### ***Características de las fuentes, sus limitaciones y el riesgo potencial de plagio***

A propósito de los argumentos anteriores, Sage (2023) presenta los siguientes puntos que prácticamente descalifican el empleo de herramientas de IA bajo sus funciones gratuitas, que posiblemente sería el uso más común que le daría la mayoría de las personas: (1) incluir la lista de fuentes usadas para generar contenidos; (2) los autores deben estar conscientes del plagio potencial que podría tener el contenido generado por IA; y (3) declarar limitaciones de las herramientas de IA, así como sus sesgos, errores y vacíos de conocimiento. Estos puntos hacen que se descarte el uso de cualquier IA en su versión gratuita, ya que las funcionalidades sin costo no permiten conocer con precisión las fuentes que está empleando la herramienta para generar el contenido que nos ofrece, así como las características de tales fuentes, lo cual sería imprescindible para poder brindar una lista de fuentes y hacer afirmaciones sobre sus limitaciones, tal como requiere Sage (2023) en los incisos 1 y 3. Estas limitaciones también incluyen lo que se conoce como el *sesgo algorítmico*, que en este contexto se refiere a aquel sesgo que existiría en las respuestas de estos sistemas, dado que las fuentes que emplea podrían presentar, incluso de manera exclusiva, a una única corriente de pensamiento, favorecer una ideología sobre otras o contener ideas discriminatorias hacia ciertos grupos; sesgos que las IA reproducirían al basar sus respuestas en tales fuentes.

En cuanto a las políticas que advierten sobre el riesgo potencial de plagio en los contenidos generados por IA, solo cuatro de las analizadas lo especifican (arXiv, 2023; ICMJE, 2023; Sage, 2023; WAME, 2023). Dadas las limitaciones de su base de conocimiento y nuestra prácticamente imposibilidad de conocer las fuentes sobre las cuales puede estar basando sus respuestas, este podría ser un riesgo bastante serio y difícil de atender sobre contenido generado por IA. Además, como *loro estocástico*, está basando lo que dice sobre lo que está en su base de conocimiento, no se trata de un ente creativo (por el momento) y si cualquiera repite las ideas contenidas en otros documentos sin incluir la apropiada atribución de la autoría de tales ideas, estamos ante una clásica definición de plagio. Este riesgo puede agravarse dado que, si bien existen herramientas para detectar textos generados por IA, también se ha dudado sobre su exactitud y confiabilidad (ver, por ejemplo, Heikkilä, 2023), aparte que llegará un momento en que podría volverse una tarea imposible, por la mejora exponencial de estas tecnologías generativas. Aparte, por si fuera poco, itambién se han ido desarrollando tecnologías para impedir la detección de contenidos generados por IA!

### ***Alteración o generación de imágenes***

Elsevier (2023c) especifica que no puede usarse IA para crear o alterar imágenes, pero que si se usa para generarlas o interpretarlas, debe ser exclusivamente en los casos en que estas herramientas sean parte del diseño de la investigación o de su metodología, para los cuales solicita que se incluya una descripción detallada del proceso empleado para generar las imágenes de tal manera que pueda reproducirse, es decir, declarar cuáles herramientas se utilizaron para su creación o alteración (versión del sistema y entidad responsable), cómo se utilizaron para tal fin e incluir la correcta atribución de tal contenido en apego a sus políticas de uso. Por su parte, dos políticas prohíben la presentación de imágenes generadas por IA (Emerald Publishing, 2023; Springer Nature, 2023), a menos que el contenido del documento trate específicamente sobre IA, lo cual será analizado caso por caso (Springer Nature, 2023).

### Consideraciones para editores

Dos consideraciones exclusivas para editores de revistas científicas que aparecen en las políticas analizadas son las siguientes: (1) Elsevier (2023b) y Sage (2023) no permiten a los editores subir manuscritos a una IA para evaluarlos, por cuestiones de confidencialidad, propiedad intelectual y también porque afirman que la evaluación puede resultar incorrecta, incompleta o sesgada y por lo tanto solo debe ser realizada por humanos. Sage (2023) y Springer Nature (2023) utilizan este mismo argumento para pedir a los revisores pares que no utilicen IA para evaluar manuscritos; y (2) por otro lado, WAME (2023) especifica que los editores necesitan tener a disposición herramientas para detectar el uso de IA sin importar que tengan o no la capacidad de pagar por estas y que se les debe ofrecer formación para emplear tales herramientas.

### Conclusión

En este trabajo he presentado algunas de las problemáticas alrededor del uso de la IA en la investigación para enmarcar la síntesis y análisis de los puntos más importantes que están presentes en diez políticas editoriales sobre el uso de IA. Una preocupación muy válida en cuanto al uso de IA en la investigación y publicación científica que podría quedarnos y que sería común a cualquier directriz que tenga que ver con la ética y la integridad científica es que estas son sugerencias, son *normas de buena conducta*, por así decirlo. El que se cumplan o no requiere en cualquier caso que la comunidad científica, así como los investigadores en formación, sean capaces de asumir una conducta íntegra y acatar las indicaciones especificadas por directrices como las aquí expuestas.

Pero a nivel ético valdría la pena preguntarse: ¿es una falta ética emplear la IA para propósitos de investigación y publicación científica? Los comentarios realizados por un foro<sup>3</sup> de COPE (2023a) incluyen la siguiente respuesta a tal pregunta:

El uso de herramientas de IA no es inherentemente poco ético y puede ser útil, por ejemplo, para autores que no escriben en inglés como su primera lengua, que tienen deficiencias de aprendizaje (por ejemplo, disgrafía) u otras dificultades al escribir. Al respecto, la IA puede ser una útil tecnología auxiliar o adaptable. El factor clave es la responsabilidad y claridad sobre su uso. La necesidad de su detección surge cuando su uso no es declarado (párr. 27).

De cualquier manera, las políticas editoriales sobre el uso de la IA son imprescindibles y es comprensible que hayan surgido como respuesta a preocupaciones alrededor de la originalidad y veracidad de los contenidos que se están presentando a las revistas y editoriales científicas. Tanta es la angustia en relación con el uso de estas tecnologías que algunos lo equiparamos al surgimiento de una especie de *plagio 2.0*, que representa un gran riesgo para la investigación y publicación científica, así como para la lectura y las habilidades de análisis y de pensamiento crítico, que podríamos decir que ya estaban mermadas en la sociedad en general antes de la aparición de estos sistemas. Prácticamente cualquiera podría estar empleando las IA de manera extensiva (no solo para la corrección de estilo u otros usos aceptables) para intentar forjarse una carrera de investigador, sin haber desarrollado y demostrado las competencias necesarias para serlo, aparte de no contar con valores de integridad, ética y honestidad científicas.

Aunque podemos estar de acuerdo con COPE en cuanto a que su uso no es esencialmente poco ético, las IA pueden dar lugar a malas prácticas. Tal como el plagio, pueden representar una forma de *hacer trampa* en las actividades científicas y pueden generarnos amplias preocupaciones en la formación de investigadores, ya que algunos podrían apostar por usar estas tecnologías en lugar de aprender a realizar por sí mismos todas las actividades de investigación, especialmente en esta era de la inmediatez, *flojera y facilismo*. Debe entenderse que la IA no puede utilizarse como la única herramienta para generar contenidos, especialmente de investigación. Es decir, lo que produzca una IA debe verse como una

<sup>3</sup> La fuente aclara que estos comentarios no implican el consenso de COPE, ni integran recomendaciones oficiales.

materia prima sobre la cual debe incorporarse el uso de otras herramientas y de la *inteligencia humana*, ya que, de no haber aportes humanos, entonces abandonemos todo, *que el último cierre la puerta*, y dejemos que las IA generen contenidos para el consumo de ellas mismas. Sin embargo, para no terminar en una nota fatalista, en el marco de las políticas analizadas y todo lo que he venido discutiendo, algunos usos aceptables de la IA en la investigación podrían incluir los siguientes:

- Trabajar con los documentos que yo seleccione utilizando las versiones de paga de los *chatbots* o los *chatbots generales especializados*, es decir, teniendo el control de la base de conocimiento a usar y realizar esta actividad solamente para análisis preliminares de grandes cantidades de texto o datos cualitativos.
- Hacer lluvias de ideas y pedir a los sistemas sugerencias de puntos de partida temáticos o posibles estudios para desarrollar una línea de investigación determinada.
- Encontrar fuentes científicas para fundamentar una investigación ya no a partir de palabras clave y búsquedas booleanas, sino de preguntas, como es el caso de Consensus y Scite. Sin embargo, entendiendo que esto no sustituye la realización de una búsqueda más amplia y sistemática en un índice bibliométrico o base de datos de mayor tamaño, que es vital para desarrollar una completa revisión de la literatura.
- Ayudar en la preparación de resúmenes y títulos, con la salvedad que es el humano quien debe revisarlos, ya que la IA puede ser muy general, poco exhaustiva y enunciativa, además que podría centrarse solo en elementos que realmente no son los más importantes o que son temas periféricos.
- Pedir a las IA que me sugieran revistas donde puedo publicar, estando conscientes que las sugerencias podrían no ser pertinentes para algunos propósitos y posibilidades individuales de publicación en cuanto a los modelos de publicación, costos o características de indizado de las revistas que recomiende.
- Apoyarse en las IA para generar contenidos de divulgación científica, con los *famosos prompts* del tipo *explícamelo como si fuera un niño de cinco años*.
- Enseñar redacción creativa, construcción y revisión de textos científicos; aparte de ética en la investigación y publicación científica.
- Procesar y trabajar con información masiva para detectar patrones iniciales que apoyen un análisis más profundo, significativo y humano.
- Detectar textos generados por IA.
- Mejorar la redacción, ortografía y gramática, además de las traducciones.
- Potenciar catálogos, repositorios, *sistemas recomendadores*, estableciendo etiquetas, patrones, relaciones, enriqueciendo descripciones y funciones de descubrimiento.

Es muy importante entender las ventajas que nos ofrecen las IA, pero también sus limitaciones. En el caso de la investigación debemos estar conscientes que no reparan en cuestiones de ética o de integridad, no se ajustan a estándares editoriales, puede ser muy difícil que citen adecuadamente y está presente el riesgo de plagio. No interpretan ni categorizan ideas a la profundidad que lo haríamos nosotros, lo cual es vital para construir, analizar y juzgar argumentos. Obviamente, tampoco las IA pueden *adivinar* lo que queremos y sus resultados podrían ser inadecuados, especialmente si no sabemos hacer consultas con un nivel de detalle suficiente, ni sabemos ser explícitos en relación a lo que queremos obtener. Por esto la construcción de *prompts* se ha vuelto un área muy importante, que incluso está representando nuevas oportunidades de trabajo para distintas áreas del conocimiento.

Después de que pase la moda de la IA, así como los tiempos de su uso y especialmente abuso, podría haber un resurgimiento de la importancia de un producto realizado al cien por ciento por humanos, así como hoy en día se aprecian mucho las artesanías y otros productos manufacturados con métodos clásicos. Características que los hacen únicos y fuera de serie.

## Referencias

- arXiv. (2023). *arXiv announces new policy on ChatGPT and similar tools*. <https://blog.arxiv.org/2023/01/31/arxiv-announces-new-policy-on-chatgpt-and-similar-tools>
- Committee on Publication Ethics. (2023a). *Artificial intelligence (AI) and fake papers*. <https://publicationethics.org/resources/forum-discussions/artificial-intelligence-fake-paper>
- Committee on Publication Ethics. (2023b). *Authorship and AI tools: COPE position statement*. <https://publicationethics.org/cope-position-statements/ai-author>
- Elsevier. (2023a). *Guide for authors*. *Journal of Biotechnology*. <https://www.elsevier.com/journals/journal-of-biotechnology/0168-1656/guide-for-authors>
- Elsevier. (2023b). *Publishing ethics*. <https://beta.elsevier.com/about/policies-and-standards/publishing-ethics>
- Elsevier. (2023c). *The use of generative AI and AI-assisted technologies in writing for Elsevier*. <https://beta.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>
- Emerald Publishing. (2023). *Publishing ethics: Find out more about publication ethics and our policies*. <https://www.emeraldgrouppublishing.com/publish-with-us/ethics-integrity/research-publishing-ethics>
- Heikkilä, M. (2023). *Why detecting AI-generated text is so difficult (and what to do about it)*. MIT Technology Review. <https://www.technologyreview.com/2023/02/07/1067928/why-detecting-ai-generated-text-is-so-difficult-and-what-to-do-about-it>
- International Committee of Medical Journal Editors. (2023). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. <https://www.icmje.org/icmje-recommendations.pdf>
- Kung, T., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., y Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Kung, T., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., y Tseng, V. (2022). *Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models*. MedRxiv. <https://doi.org/10.1101/2022.12.19.22283643>
- Machin-Mastromatteo, J. D. [Juanfíco]. (20 de abril de 2022). *¿Spam en Google Académico?* [Video]. YouTube. <https://youtu.be/1lN1RoaV4BU>
- Orduña Malea, E. [@eomalea]. (24 de abril de 2023). *I'm afraid to say that several preprint servers are publishing online papers, which cite publications co-authored by me that do not exist. This is the @chatgptimpact. Google Scholar and ResearchGate are indexing those papers, and their fake citations, by the way.* [Tweet]. Twitter. <https://twitter.com/eomalea/status/1650527418577309699>
- Oxford University Press. (2023). *Ethics*. Oxford Academic. [https://academic.oup.com/pages/authoring/journals/preparing\\_your\\_manuscript/ethics](https://academic.oup.com/pages/authoring/journals/preparing_your_manuscript/ethics)
- Sage. (2023). *ChatGPT and generative AI: Use of large language models and generative AI tools in writing your submission*. <https://us.sagepub.com/en-us/nam/chatgpt-and-generative-ai>

- Spinak, E. (2023). *¿Es que la Inteligencia Artificial tiene alucinaciones?* SciELO en Perspectiva. <https://blog.scielo.org/es/2023/12/20/es-que-la-inteligencia-artificial-tiene-alucinaciones>
- Springer Nature. (2023). *Artificial intelligence (AI)*. Nature Portfolio. <https://www.nature.com/nature-portfolio/editorial-policies/ai>
- Tang, G., y Eaton, S. (2023). *A rapid investigation of artificial intelligence generated content footprints in scholarly publications*. Research Square. <https://doi.org/10.21203/rs.3.rs-3253789/v1>
- Taylor & Francis Group. (2023). *Defining authorship in your research paper: Co-authors, corresponding authors, and affiliations*. Author Services. <https://authorservices.taylorandfrancis.com/editorial-policies/defining-authorship-research-paper>
- Tsai, C., Yeh, Y., Tsai, L., y Chou, E. (2023). The efficacy of transvaginal ultrasound-guided BoNT-A external sphincter injection in female patients with underactive bladder. *Toxins*, 15(3), 199. <http://doi.org/10.3390/toxins15030199>
- Villegas-Ceballos, S. [Santiago Villegas-Ceballos]. (2 de diciembre de 2023). *Inteligencia Artificial en Bibliotecas - Introducción 2023-12* [Video]. YouTube. <https://www.youtube.com/watch?v= klpXNc7vKw>
- Wiley. (2023). *Best practice guidelines on research integrity and publishing ethics*. <https://authorservices.wiley.com/ethics-guidelines/index.html>
- World Association of Medical Editors. (2023). *Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications*. <https://wame.org/page3.php?id=106>