

¿Es la inteligencia artificial generativa una aliada de la ciencia abierta?

Is generative artificial intelligence an ally of open science?

Alexandre López-Borrull

López-Borrull, Alexandre (2024). "¿Es la inteligencia artificial generativa una aliada de la ciencia abierta?". *Anuario ThinkEPI*, v. 18, e18e40.

<https://doi.org/10.3145/thinkepi.2024.e18a40>

Publicado en *IweTel* el 17 de diciembre de 2024

Alexandre López-Borrull

<https://www.directorioexit.info/ficha157>

<https://orcid.org/0000-0003-1609-2088>

Universitat Oberta de Catalunya

alopezbo@uoc.edu



Resumen: El artículo explora la relación entre la inteligencia artificial generativa (IAG) y la ciencia abierta, destacando tanto sus beneficios como sus riesgos. La IAG puede simplificar conceptos científicos complejos, mejorar la gestión de datos y automatizar procesos editoriales, lo que podría potenciar la ciencia abierta. Sin embargo, también presenta desafíos éticos y riesgos, como la generación de síntesis erróneas, la perpetuación de sesgos y la creación de contenido de baja calidad. La relación entre la IAG y la ciencia abierta se compara con interacciones biológicas como el mutualismo y el parasitismo, sugiriendo que un enfoque equilibrado y colaborativo es esencial. La ética y la transparencia son fundamentales para asegurar que los datos abiertos se utilicen de manera responsable y que los resultados generados sean verificables y reproducibles.

Palabras clave: Inteligencia artificial generativa; Ciencia abierta; Ética; Datos de investigación.

Abstract: The article explores the relationship between generative artificial intelligence (GenAI) and open science, highlighting both its benefits and risks. GenAI can simplify complex scientific concepts, improve data management, and automate editorial processes, which could enhance open science. However, it also presents ethical challenges and risks, such as generating incorrect summaries, perpetuating biases, and creating low-quality content. The relationship between GenAI and open science is compared to biological interactions like mutualism and parasitism, suggesting that a balanced and collaborative approach is essential. Ethics and transparency are fundamental to ensure that open data is used responsibly and that the generated results are verifiable and reproducible.

Keywords: Generative artificial intelligence; Open science; Ethics; Research data.

Introducción

La especialización en nuestros campos científicos a menudo nos lleva a crear compartimentos estancos que separan nuestra realidad de nuestros contextos. Ante un punto de inflexión significativo, nos resulta difícil integrar este nuevo elemento en nuestro camino debido al desconocimiento o la inercia. Esto es particularmente evidente en el caso de la inteligencia artificial generativa (IAG), que se explora más como una herramienta que como un motor de cambio conceptual.

Este otoño intenté desafiar al asistente digital *Copilot* sobre la relación entre la ciencia abierta y la IAG. Reflexionando sobre las interacciones entre especies (mutualismo, parasitismo, depredación...), me pregunté cuál de estas relaciones podría aplicarse en este caso. Le pregunté específicamente quién se beneficiaba más de esta intersección, si la IAG o la ciencia abierta. Su respuesta, que inició la reflexión de este artículo, fue clara: “La intersección entre la IAG y la ciencia abierta ofrece beneficios significativos tanto para la IAG como para la ciencia abierta, pero en términos de quién se beneficia más, parece que la ciencia abierta obtiene las ventajas más sustanciales”. ¿Es esto cierto o estamos pasando por alto algo en el camino?

En la revisión de la poca literatura existente sobre los desafíos y oportunidades que la IAG presenta para la ciencia abierta, se destacan algunas ideas. En primer lugar, muy recientemente **Hosseini et al.** (2024) exploran tanto los impactos positivos como negativos de la IAG en las prácticas de la ciencia abierta. Los autores subrayan que, aunque puede simplificar conceptos científicos complejos y mejorar la gestión de datos y la programación, la IAG también puede generar síntesis erróneas (las famosas alucinaciones), además de introducir inexactitudes en el código y producir informes de revisión superficiales, por lo que hace énfasis en la baja confianza en la creación de contenidos por parte de la IAG. Así, apuntan que, aunque tiene el potencial de avanzar en objetivos compartidos con la ciencia abierta, también presenta riesgos significativos que deben ser considerados para no caer en la fe ciega de la nueva tecnología. Como en otros ámbitos, los autores apuntan a un escenario abierto con ventajas e inconvenientes por explorar.

Por otro lado, **Acion et al.** (2023), se habían centrado en los desafíos éticos que la IAG planteaba para la ciencia abierta. Los autores argumentan que la IAG puede causar discriminación y daños, contraviniendo los propios valores del movimiento por la ciencia abierta. Además, señalaron que la ciencia abierta proporciona datos abiertos de alta calidad que son utilizados por la IAG, lo que plantea el dilema de cómo compartir datos sin que sean utilizados de manera perjudicial. Así, proponen crear mecanismos de gobernanza para proteger el conocimiento como bien común y el derecho a la investigación como un derecho humano. Es interesante cómo la propia alimentación de la herramienta hace reflexionar sobre el público potencial de la compartición de artículos o datos, como algo que no se hubiera tenido en cuenta hasta ahora o se pudiera conectar o desconectar fácilmente, como un interruptor. Trataremos este asunto más adelante.

Finalmente, en un enfoque más práctico, **Zhou** (2024) aborda cómo la IAG puede ayudar a los editores a enfrentar sus desafíos en la transformación digital y el movimiento de la ciencia abierta. En este sentido, se destaca la utilización de la IAG para mejorar la visibilidad y accesibilidad de los contenidos, automatizar procesos editoriales, detectar problemas de integridad en la investigación, así como para diversificar las fuentes de ingresos. Como resumen, mientras algunos autores consideran la IAG como una herramienta para mejorar y hacer más eficiente la compartición de conocimiento, otros autores adoptan una perspectiva crítica sobre los riesgos éticos.

Ventajas e inconvenientes del uso de la IAG en ciencia abierta

Como comentábamos anteriormente, la relación entre estos dos conceptos puede compararse con las interacciones biológicas entre especies, como el parasitismo y el mutualismo. En un escenario de parasitismo, la IAG podría aprovecharse de los datos abiertos sin contribuir de vuelta, beneficiándose unilateralmente. Por otro lado, en un escenario de mutualismo, tanto la IAG como la ciencia abierta se benefician mutuamente: la IAG mejoraría gracias a los datos accesibles, y la ciencia abierta se enriquecería con nuevas herramientas y nuevos conocimientos facilitados por la IAG. Esta comparación ilustra la complejidad y el potencial de la interacción entre la IAG y la ciencia abierta, destacando la necesidad de un enfoque equilibrado y colaborativo. ¿No debíamos compartir

sin filtro, no era eso la ciencia abierta? Sin duda, el punto crítico para que esta relación sea verdaderamente beneficiosa es que la IAG se desarrolle y utilice de manera ética y transparente, asegurando que los datos abiertos sean utilizados de manera responsable y que los resultados generados por la IAG sean verificables y reproducibles. Y esto, aunque no está asegurado, no debería ser según mi punto de vista, un motivo para cerrar contenidos sino para pedir más garantías y reflexión. Y más en un contexto donde la existencia de contenidos de calidad es básica en todos los entornos, ya sean sitios web o redes sociales, pero también repositorios académicos y servidores de *preprints*.

la introducción de la IAG puede beneficiar al conocimiento científico si ayuda a mejorar su calidad, no su cantidad

Sin duda, el beneficio mutuo pasaría por entender que la introducción de la IAG puede beneficiar al conocimiento científico si ayuda a mejorar su calidad, no su cantidad. Si consideramos que la IAG nos va a ayudar a crear más contenidos, más artículos, más *datasets*..., tendremos mucha más ciencia abierta, pero no necesariamente mejor ciencia. Y aquí es donde se vislumbra el punto más importante. La ciencia abierta no es un fin por sí mismo sino un medio para mejorar el proceso de diseminación del conocimiento y por ende la propia ciencia. ¿Deberían las herramientas de IAG beber de repositorios y revistas científicas? Desde esta visión, sin duda. ¿Pueden existir riesgos éticos y legales? Sí, aquellos intrínsecos a la propia herramienta (sesgos, desinformación, incumplimiento de los derechos de autor), pero parece más una condición para el trabajo conjunto que para crear falsas barreras que pueden ser derivadas (vía *Scihub*, por ejemplo).

A nivel práctico, la IAG puede mejorar sin duda la gestión y el uso de datos de investigación abiertos. Herramientas basadas en IAG pueden ayudar en la creación de planes de gestión de datos, la validación y limpieza de datos, y la generación de metadatos. Esto facilita la reutilización y la interoperabilidad de los datos, siendo más FAIR que nunca. Al automatizar estos procesos, la IAG permite a los investigadores centrarse en tareas más creativas, aumentando la eficiencia y la productividad en la investigación científica. De nuevo, pero, esta eficiencia debe enfocarse en tener incentivos para mejorar los datos y filtrar aquellos que pueden aportar valor de los que no.

Sin embargo, todo lo expuesto anteriormente no omite que haya que considerar también otro posible aspecto negativo, que sería la posibilidad de que la IAG perpetúe y amplifique los sesgos existentes en los datos de entrenamiento. Si los modelos de IAG se entrenan con datos que contienen sesgos de género, raza o socioeconómicos, estos sesgos pueden reflejarse y amplificarse en los resultados generados. Nuestros sesgos creando nuevos sesgos. Esto no solo contraviene los principios de equidad y diversidad de la ciencia abierta, sino que también puede llevar a la exclusión de ciertos grupos y perspectivas en la investigación científica. Además, la falta de transparencia en los algoritmos y procesos de la IAG puede dificultar la identificación y corrección de estos sesgos.

Asimismo, debemos también considerar que la IAG puede facilitar la creación de contenido científico de baja calidad o incluso fraudulento. La capacidad de generar textos y artículos científicos de manera automática puede ser explotada por actores malintencionados para producir publicaciones falsas o de baja calidad, lo que aumenta el ruido en la literatura científica y dificulta la identificación de investigaciones válidas y relevantes. Esto puede afectar negativamente la visibilidad y la credibilidad de la ciencia abierta, ya que la proliferación de contenido de baja calidad puede desincentivar la participación y la colaboración en la comunidad científica. En resumen, aunque la IAG tiene el potencial de beneficiar a la ciencia abierta, también presenta riesgos significativos que deben ser gestionados cuidadosamente para evitar daños al movimiento.

Como hemos comentado anteriormente, la apertura de contenidos en el ámbito de la ciencia abierta mediante la IAG plantea importantes consideraciones éticas. En el caso de la IAG, la disponibilidad de grandes repositorios de datos puede mejorar la precisión y creatividad de los modelos, pero también plantea preguntas sobre la privacidad y el consentimiento de los datos utilizados. Es esencial establecer directrices claras sobre el uso y la distribución de datos en repositorios y revistas científicas, así como fomentar una cultura de responsabilidad y ética en la investigación.

La ética en la ciencia abierta y la IAG también implica considerar quién tiene acceso a estos recursos y cómo se utilizan. La idea de abrir contenidos a todo el mundo, pero no necesariamente a todas las inteligencias, plantea un dilema sobre la equidad y la justicia en la distribución del conocimiento. Si bien la ciencia abierta busca democratizar el acceso a la información, es crucial asegurarse de que este acceso no se limite a aquellos con las capacidades tecnológicas para aprovecharlo plenamente. Además, es importante reflexionar sobre cómo se pueden proteger los derechos de los individuos y las comunidades cuyos datos se utilizan en la investigación, garantizando que se respeten su privacidad y autonomía.

Conclusión

La aceleración tecnológica que implica la IAG tiene la capacidad de transformar el panorama científico. La tecnología permite la creación de plataformas más eficientes para compartir y acceder a datos científicos, lo que puede potenciar la colaboración y la innovación. Sin embargo, esta rápida evolución también puede generar una sobrecarga de información, creando una “burbuja de ruido documental” que dificulte la identificación de información relevante y de calidad. Parece oportuno en este sentido desarrollar estrategias para gestionar este flujo de datos y garantizar que la ciencia abierta mantenga su objetivo de mejorar la calidad y accesibilidad de la investigación científica. La integración de la IAG en estos procesos puede ser una herramienta valiosa, siempre y cuando se utilice de manera ética y responsable, preservando el espíritu original de la ciencia abierta: hacer la ciencia más accesible y colaborativa, no simplemente aumentar la cantidad de información disponible. Dicha aceleración y ahorro de tiempo no debería usarse para generar más contenidos, sino mejor conocimiento.

En conclusión, la relación entre la ciencia abierta y la IAG como se ha visto puede definirse como incipiente, compleja, llena de potencial, matices y riesgos. La IAG tiene el potencial de transformar la manera en que se produce y se comparte el conocimiento, pero es crucial abordar las consideraciones éticas y gestionar adecuadamente la aceleración tecnológica para asegurar que estos avances beneficien adecuadamente a la comunidad científica y a la sociedad en general. La ciencia abierta debe seguir promoviendo la transparencia, la colaboración y la accesibilidad, integrando de manera responsable cualquier nueva tecnología para cumplir con su misión de mejorar la ciencia, para todos, para siempre.

Referencias

Acion, Laura; Rajngewerc, Mariela; Randall, Gregory; Etcheverry, Lorena (2023). “Generative AI poses ethical challenges for open science”. *Nature Human Behaviour*, n. 7, pp. 1800–1801.

<https://doi.org/10.1038/s41562-023-01740-4>

Hosseini, Mohammad; Horbach, Serge P. J. M.; Holmes, Kristi; Ross-Hellauer, Tony (2024). “Open science at the generative AI turn: An exploratory analysis of challenges and opportunities”. *Quantitative science studies*.

https://doi.org/10.1162/qss_a_00337

Zhou, Hong (2024). “The top ten challenges, needs, and goals of publishers – and how AI can help in digital transformation and the open science movement”. *The Scholarly Kitchen*, 5 November.

<https://scholarlykitchen.sspnet.org/2024/11/05/the-top-ten-challenges-needs-and-goals-of-publishers-and-how-ai-can-help-in-digital-transformation-and-the-open-science-movement>