

Asistentes de IAGen en la investigación académica: revolucionando la recuperación y análisis de literatura científica

GenAI Assistants in academic research: revolutionizing scientific literature retrieval

ROBERTO SANTOS SOLÓRZANO,* DAFNE ESTEFANY BONILLA CERQUEDO**

* Coordinación de Evaluación, Innovación y Desarrollo Educativos (CEIDE). Universidad Nacional Autónoma de México. Circuito, Mario de La Cueva 20, Insurgentes Cuicuilco, Coyoacán, Ciudad de México, México. Correo electrónico: roberto_santos@ceide.unam.mx. ORCID: <https://orcid.org/0000-0002-2469-0048>

** Coordinación de Evaluación, Innovación y Desarrollo Educativos (CEIDE). Universidad Nacional Autónoma de México. Circuito, Mario de La Cueva 20, Insurgentes Cuicuilco, Coyoacán, Ciudad de México, México. Correo electrónico: dafne_bonilla@ceide.unam.mx. ORCID: <https://orcid.org/0000-0002-1486-1587>

Biblioteca Universitaria, vol. 27, núm 2, julio–diciembre 2024, pp. 109–120.
DOI: <https://doi.org/10.22201/dgbsdi.0187750xp.2024.2.1554>

Palabras Clave:

IAGen, asistentes de IA, inteligencia artificial, investigación, literacidad digital, recuperación de literatura.

Keywords:

Assistants AI, artificial intelligence, digital literacy, GenAI, investigation, literature retrieval.

RESUMEN

La inteligencia artificial (IA) es un campo interdisciplinario que, en los últimos dos años, ha transformado radicalmente la sociedad, el trabajo, la comunicación, la educación y nuestra manera de relacionarnos. Su impacto ha facilitado la automatización y conectividad global en una escala sin precedentes. Particularmente, la inteligencia artificial generativa (IAGen) y los asistentes para la recuperación de literatura científica han revolucionado la investigación académica. Estas herramientas no sólo simplifican el proceso de búsqueda y la recopilación de literatura, sino que también atienden las críticas sobre alucinación y la limitación para proporcionar fuentes adecuadas, problemas frecuentes en otras herramientas de IAGen, como los chats.

El campo de la IAGen y los asistentes se encuentran en constante cambio y mejora, pero su uso plantea importantes desafíos, como: la falta de un marco regulatorio, la opacidad en los algoritmos, la posible amplificación de sesgos y la necesidad de desarrollar la literacidad digital en los usuarios para garantizar un uso informado, responsable y ético. Este trabajo examina críticamente tanto los beneficios como las limitaciones de estas herramientas, y propone la literacidad digital como una competencia esencial que se debe desarrollar en el contexto académico.

ABSTRACT

Artificial intelligence (AI) is an interdisciplinary field that, over the past two years, has radically transformed society, employment, communication, education, and the ways in which we interact. Its impact has facilitated automation and global connectivity on an unprecedented scale. In particular, generative artificial intelligence (GenAI) and scientific literature retrieval assistants have revolutionized academic research. These tools not only simplify the process of searching for and collecting literature but also address criticisms regarding hallucinations and limitations in providing reliable sources—challenges commonly found in other GenAI tools, such as chatbots.

The fields of GenAI and assistant technologies are constantly evolving and improving. However, their use presents significant challenges, including the lack of regulatory frameworks, algorithmic opacity, the potential amplification of biases, and the need to foster digital literacy among users to ensure informed, responsible, and ethical use. This paper critically examines the benefits and limitations of these tools and argues that digital literacy is an essential competency that should be developed in the academic context.

Introducción

La inteligencia artificial (IA) es un campo interdisciplinario que ha evolucionado significativamente desde su origen, hace aproximadamente 70 años. Esta combina elementos de distintas disciplinas para desarrollar sistemas capaces de realizar tareas que, hasta hace poco, se consideraban propias de la inteligencia humana. La IA se conceptualiza de diversas maneras, dependiendo del contexto y la perspectiva adoptada. Algunos la consideran un campo de estudio, otros un conjunto de habilidades, y otros más como sistemas autónomos capaces de percepción, comprensión, acción y aprendizaje (Perkins y Roe, 2023).

Estas perspectivas reflejan –por un lado– la diversidad y complejidad del campo y –por otro– el hecho que abarca diferentes dimensiones. Por ello encontramos un sinnúmero de definiciones, desde muy generales, como la que propone la Comisión Europea, que desde 2018 señala que la IA se refiere a los sistemas que muestran un comportamiento inteligente al analizar su entorno y tomar medidas, con cierto grado de autonomía, para lograr objetivos específicos (Boucher, 2020).

Pero, en paralelo se plantean definiciones más detalladas como la de Martínez (2024), que señala que la inteligencia artificial es un campo que surge de la interacción entre disciplinas, como: la neurociencia, la

psicología cognitiva, la matemática, la biología evolutiva, las ciencias de la computación y la lógica. Añade que, aunque no exista una única definición, la IA se enfoca a estudiar procesos observados en la naturaleza que pueden clasificarse como inteligentes (la adquisición y el empleo del lenguaje natural, las formas para resolver problemas y la manera en que se estructura el conocimiento), con el objetivo de reproducir estos esquemas en un programa de computadora o en algoritmos.

Entre los distintos tipos y clasificaciones de IA resalta la denominada IA Generativa (IAGen), que combina la programación informática con grandes conjuntos de datos para generar contenidos como texto, imagen, video, música y código, de manera automática a partir de *prompts* o instrucciones escritas (UNESCO, 2024). La IAGen utiliza modelos de procesamiento de lenguaje natural avanzado (NLP, *Natural Language Processing*) y grandes modelos de lenguaje (LLMs, *Large Language Models*); por ejemplo, los modelos como GPT-4 (*Generative Pre-trained Transformer 4*) requieren una enorme cantidad de datos y fases de entrenamiento para identificar patrones hasta que pueda reconocerlos y realizar una gran variedad de tareas (Bender, *et al.*, 2021; Mostafapour, *et al.*, 2024).

Particularmente con esta tecnología existe un creciente interés por su potencial para desarrollar herramientas y aplicaciones que puedan asistir en tareas

académicas, incluida la recuperación y revisión de literatura; estas herramientas han demostrado una enorme eficacia en las revisiones generadas en comparación con los métodos tradicionales, debido a que es posible extraer información clave de grandes cantidades de literatura en poco tiempo, además que permiten la identificación automática de hipótesis, métodos y conclusiones en los artículos, así como la categorización de tendencias emergentes dentro de un campo.

No obstante, el campo de la IAGen enfrenta varios desafíos, especialmente en términos de: establecer marcos de gobernanza que aseguren su uso ético y sostenible, la posible falta de transparencia en cómo seleccionan y presentan la información, la necesidad de ajustar los algoritmos para evitar sesgos en los resultados (Barsha y Munshi, 2024). Otro desafío, pero a nivel de usuario, es que las personas requieren desarrollar la literacidad digital, concepto que va más allá de la alfabetización digital, ya que no solo se trata de saber usar estas herramientas, sino también de entender y cuestionar cómo y por qué funcionan de cierta manera (Abio, 2017; Jones y Hafner, 2021; Méndez–Ochaita, 2022).

En el contexto de la IAGen la literacidad digital es crucial, ya que las respuestas que se generan las construye a partir de patrones y datos, muchas veces sin que los usuarios sepan exactamente cómo o con qué criterios. Ante este panorama, el objetivo de este trabajo es explorar críticamente el impacto y las oportunidades que ofrecen los asistentes de IAGen para la recuperación y el análisis de literatura en la investigación académica.

La era digital y la (r)evolución de la información

Los desarrollos del campo de la recuperación de literatura y el de la inteligencia artificial están profundamente ligados con la era digital y la (r)evolución de la información. La primera se refiere al periodo que inició con la integración de las computadoras en la práctica cotidiana durante la segunda mitad del siglo XX, lo que condujo al rápido desarrollo y la adopción de Internet y la aparición de una gran variedad de dispositivos electrónicos (Trujano *et al.*, 2008). Por su parte, la (r)

evolución de la información, se refiere a los cambios drásticos en la forma y los medios por los cuales se genera, recolecta, almacena y transmite la información (Michán, Álvarez y Guillén, 2013, Maldonado, 2020).

Estos fenómenos provocaron, entre otras cosas, el desarrollo de una gran cantidad y diversidad de recursos, herramientas y aplicaciones digitales (Michán, Álvarez y Guillén, 2013, Maldonado, 2020), un crecimiento tecnológico sin precedentes, una mejora en la calidad de los servicios informacionales en todas las áreas del conocimiento (Trujano *et al.*, 2008, Maldonado, 2020), pero sobre todo un aumento colosal en la cantidad y diversidad de información, a tal punto que la cantidad de datos digitales creados para 2024 a nivel mundial llegó a 147 zettabytes (Statista, 2024).

La era digital y la (r)evolución de la información actualmente abarcan desarrollos tecnológicos como la IA, el *blockchain* y el Internet de las cosas (IoT), propuestas que van más allá del manejo de información, dado que están transformando radicalmente la sociedad, el trabajo, la comunicación, la educación y la forma en que nos relacionamos, facilitando la automatización y conectividad global en una escala sin precedentes (Carbonell–García, 2023).

En el ámbito académico trajeron una avalancha en la producción científica tanto de artículos como de revistas, lo que consecuentemente provoca que haya una vasta cantidad de información sobre prácticamente cualquier tema, pero en contraparte da lugar a una sobrecarga de información (Rzhetsky *et al.*, 2008); por ello, surgen constantemente nuevos y mejores métodos para la búsqueda y análisis de toda esta literatura, como: la generación de grandes colecciones bibliográficas y bases de datos digitales, el desarrollo de un gran número de recursos electrónicos como buscadores generales, semánticos y basados en IA, así como la generación de sistemas de información más robustos (Egger y Carpi, 2009).

La consulta de literatura científica tiene un alto valor entre las disciplinas, ya que permite a los académicos apropiarse de nuevos conocimientos, comparar entre distintos puntos de vista, establecer relaciones,

realizar análisis críticos, plantear hipótesis, formular nuevas preguntas, extraer información y enriquecer su vocabulario; no obstante, por la propia naturaleza, estructura y propósito de estos textos, puede resultar difícil el recuperarlos (Egger y Capri, 2009). Esta nueva escena informacional e informática implica un perfil más activo, donde se requiere conocer, entender, usar y desarrollar nuevas estrategias, términos y métodos para enfrentarse a los nuevos paradigmas tecnológicos (Jardón *et al.*, 2024).

Sin embargo, la tarea de recuperar y analizar toda la literatura es un enorme desafío debido a la cantidad exponencial de investigaciones que se publican cada día, al acceso restringido a una gran cantidad de revistas científicas, al desconocimiento o poco uso de los recursos informáticos e informacionales disponibles en las instituciones y a la escasa formación y desarrollo de habilidades en la recuperación de literatura (Jardón *et al.*, 2024).

Las herramientas de IA están comenzando a jugar un papel crucial en este campo. Sin embargo, estos sistemas no están exentos de discusiones, como la posible falta de transparencia y la necesidad de ajustar los algoritmos para evitar sesgos en los resultados (Barsha y Munshi, 2024).

La recuperación de literatura. Del método tradicional a la IAGen

La recuperación de literatura (*literature retrieval* en inglés) es un concepto muy amplio y complejo, que incluso abarca la descripción de toda una disciplina, pero para los fines de este trabajo ocuparemos la propuesta de la Universidad Finlandesa de Aalto (2023), que se refiere a ella como un proceso sistemático que trata sobre la búsqueda de literatura digital en cualquier tipo de colección documental electrónica, a través de Internet, con el fin de poderla usar de manera pertinente (Aalto University Learning Centre, 2023). Aunque parece una sola actividad, se trata de una serie de tareas organizadas que no solo pretenden buscar, localizar y acceder a la literatura disponible en Internet, sino también evaluarla, seleccionarla y gestionarla para resolver las necesidades de información (Maldonado, 2020).

La recuperación de literatura permite: identificar una amplia gama de referencias sobre un tema específico, extraer información para elaborar directrices basadas en evidencia, profundizar en el método de investigación y la evaluación académica, además, para identificar el diseño del estudio, la población muestreada, determinar las variables, analizar resultados e identificar posibles fallas y cómo evitarlas (Grewal, Kataria y Dhawan, 2016).

Se trata de una actividad interdisciplinaria que se relaciona con la bibliotecología, la informática y las ciencias de la información. En consecuencia, esto provoca que los académicos desarrollen un conocimiento parcial o limitado sobre lo que es y cómo realizarla. De este modo, es importante la formación en habilidades transversales desde distintos enfoques y el estar informados sobre la extensa variedad de fuentes de información y herramientas pertinentes (Salvador, *et al.*, 2006; Michán, Álvarez y Guillén, 2013).

De manera tradicional existen varios métodos de búsqueda bibliográfica, durante muchas décadas, se realizaba directamente en las bibliotecas revisando la literatura en formato físico. Con el avance de la tecnología, principalmente de Internet, la recuperación de literatura se realiza en motores de búsqueda basados en la web o en bases de datos bibliográficas (Salvador *et al.*, 2006; Grewal, Kataria y Dhawan, 2016).

Estas estrategias implican que se construya una consulta a partir de palabras clave para expresar la necesidad de información, pero este método difícilmente captura toda la complejidad, entre otras razones, porque las palabras clave son a menudo imprecisas, incompletas, ambiguas e inadecuadas; además, la propia consulta impone un conjunto de restricciones al proceso de recuperación.

Finalmente, por más sofisticadas que sean las bases de datos, ninguna puede encontrar toda la literatura, por lo que es necesario realizar el ejercicio de recuperación en varias colecciones diferentes, lo que conlleva una cantidad considerable de tiempo y esfuerzo en la búsqueda de información, para navegar por bases de datos, revisar índices y realizar búsquedas



en catálogos de bibliotecas, por lo que este proceso, aunque metódico, se encuentra limitado por la capacidad humana para procesar grandes volúmenes de información (Boughanem, 2020).

Las herramientas, aplicaciones y asistentes de IAGen favorecen considerablemente la tarea de recuperación de literatura, al abordar la incompletitud, la vaguedad y la subjetividad intrínsecas del proceso manual. Los modelos de IA se introdujeron en los sistemas de recuperación de información en la década de 1980, transformando los sistemas clásicos basados en modelos booleanos hacia sistemas basados en la clasificación y enfoques probabilísticos. Hoy en día, se han desarrollado varios enfoques de recuperación que incluyen técnicas de IA que ayudan a expresar el contenido del documento, a comprender las necesidades de información y a formalizar eficazmente el concepto de relevancia (Boughanem, 2020).

IAGen y su impacto en la investigación académica

La incorporación de herramientas y aplicaciones de IAGen en el ámbito académico no solo está transformado los procesos de búsqueda y análisis de información, sino que también ha redefinido significativamente el papel de

los involucrados que ahora requieren del desarrollo de la literacidad digital, concepto que define nuevas habilidades y competencias para utilizar los asistentes de IAGen de manera efectiva y ética, permitiéndoles interpretar y evaluar críticamente la relevancia y la confiabilidad de los resultados que estas producen (Méndez-Ochaita, 2022; Van Noorden y Perkel, 2023).

La literacidad digital es esencial para profundizar tanto en el conocimiento sobre el funcionamiento de los algoritmos, como en la comprensión de las limitaciones y posibles sesgos de estas herramientas (Van Noorden y Perkel, 2023). Hacer un uso informado y responsable, sumado al desarrollo de las habilidades de pensamiento crítico, es indispensable para mantener los estándares de rigor académico en un entorno cada vez más dominado por estas tecnologías, por lo que resulta inherente el conocimiento sobre las implicaciones éticas, cuestiones de privacidad, propiedad intelectual y de reproducibilidad (Méndez-Ochaita, 2022; Fernández-Samos-Gutiérrez, 2023; Van Noorden y Perkel, 2023).

Ante este cambio de paradigma de los métodos de recuperación de literatura, y a medida que la IAGen evoluciona, es crucial que la comunidad académica participe activamente en la discusión de cómo estas tecnologías se integran en la práctica de la

investigación. En la actualidad, los académicos se encuentran en la transición de consumidores de información a co-creadores de conocimiento junto con los sistemas de IAGen, cada vez que estas herramientas pueden sugerir hipótesis, identificar brechas en la investigación y proponer nuevas direcciones de estudio. De tal modo que el trabajo de investigación está sujeto a la colaboración, a la interacción y a la iteración constante con estos sistemas para refinar sus preguntas de investigación y metodologías (Perkins y Roe, 2024).

Ventajas y alcances de los asistentes de IAGen para la recuperación

Es importante resaltar que estos asistentes, a diferencia de otras herramientas de IAGen populares o de uso general como los chats, están especializados en encontrar, resumir y analizar publicaciones académicas; se han desarrollado para realizar búsquedas en bases de datos bibliográficas validadas como Semantic Scholar, Open Alex, Crossref, entre otras, con diversos modelos de recuperación y tipos de acceso (ANEXO 1).

Estos asistentes ofrecen diversas ventajas, como: la de optimizar la precisión en las búsquedas y su eficiencia, transformar el proceso de la investigación académica tradicional, analizar grandes volúmenes de datos, obtener resultados categorizados e indexados, ahorro significativo de tiempo en el diseño de las consultas, a la vez que amplían considerablemente el alcance de la literatura recuperada (Mostafapour, *et al.*, 2024; Van Noorden y Perkel, 2023).

Una ventaja considerable en la investigación académica, es que tienen como fuente de información bases de datos especializadas que almacenan y organizan publicaciones científicas validadas y revisadas por pares, reduciendo el margen de “alucinaciones”, concepto que se refiere a situaciones en las que el modelo genera información que parece plausible pero que en realidad es inexacta, inventada o carece de fundamento en los datos reales; esto además ayuda a subsanar la crítica de que los modelos de IAGen no están diseñados para proporcionar citas o referencias verificables (Bender, *et al.*, 2021; Ji, *et al.*, 2023).

Estas tecnologías también ofrecen la capacidad para detectar patrones y conexiones que podrían pasar desapercibidos en una búsqueda convencional, ya que pueden llevar a cabo análisis bibliométricos de mayor complejidad, como la identificación y formación que redes, es decir, que permiten observar y mapear relaciones complejas entre diferentes campos de investigación, autores y conceptos, lo cual enriquece la comprensión de las dinámicas de la producción científica y visibiliza el impacto de diferentes trabajos y autores en una determinada comunidad (Bender, *et al.*, 2021; Van Noorden y Perkel, 2023).

Este impacto se traduce a la vez en una mejora de la accesibilidad a la literatura académica; por ejemplo, la traducción automática avanzada es un gran apoyo al navegar por documentos con contenidos complejos y para superar barreras lingüísticas en la investigación académica. Así, el acceso a literatura en múltiples idiomas amplía el alcance global de las investigaciones (UNESCO, 2024; Mostafapour, *et al.*, 2024; Patel, *et al.*, 2024; Perkins y Roe, 2024).

Finalmente, los asistentes de IAGen están facilitando la detección de plagio y la verificación de la integridad académica. Los algoritmos avanzados pueden comparar rápidamente nuevos manuscritos con una robusta base de datos de publicaciones existentes, identificando similitudes sospechosas y ayudando a mantener los altos estándares de originalidad en la investigación académica (Mostafapour, *et al.*, 2024; Perkins 2023; Perkins y Roe, 2023).

Es un hecho que el uso de estos asistentes ha revolucionado el acceso y análisis de información en el ámbito académico. Sin embargo, a pesar de sus innegables beneficios, la implementación de esta tecnología aún enfrenta desafíos, limitaciones y preocupaciones tanto éticas como técnicas que surgen de su aplicación.

Desventajas y limitaciones de los asistentes de IAGen para la recuperación

Existen múltiples señalamientos respecto al uso de estos asistentes, ya que mucho del conocimiento de cómo funcionan es insuficiente, e incluso es posible

que esté influenciado por prejuicios. Los LLMs que forman la base de muchas aplicaciones de IAGen perpetúan y amplifican sesgos presentes en los grandes volúmenes de datos con los que se entrenan, ya que no están exentos de ser *de baja calidad, estar incompletos o presentar ciertas tendencias, lo que lleva a respuestas parciales o discriminatorias, por ejemplo, la preferencia por ciertas fuentes o autores, el idioma, o hasta la subrepresentación de perspectivas minoritarias o trabajos de regiones menos representadas en las bases de datos digitales* (Campolo, *et al.*, 2017; UNESCO, 2024).

A esto se añade la falta de transparencia en los algoritmos y la opacidad en la generación de los modelos de “caja negra” utilizados en muchos sistemas, lo que implica que los usuarios no puedan comprender del todo cómo y por qué un algoritmo toma ciertas decisiones y genera ciertas respuestas, o cómo impacta en la privacidad y protección de sus datos. Esto representa un desafío significativo para determinar responsabilidades y garantizar la reproducibilidad en la investigación académica, lo que también abona a la inequidad (Campolo, *et al.*, 2017; Bender, *et al.*, 2021; Guo y Caliskan, 2021). Esta falta de transparencia en los LLMs puede deberse, entre otras causas, a que muchos de estos asistentes son desarrollados por empresas que se reservan el derecho a resguardar sus datos, ya que es parte de estrategias de preservación y depuración de información confidencial (Novelli, C., 2024).

A esto también se añade la tendencia de los asistentes de IAGen a perder sutilezas y contextos importantes en el análisis de textos académicos, es decir, que puede existir una interpretación errónea o que se simplifique excesivamente, lo cual puede traer como consecuencia la recuperación y la propagación de información inexacta o incompleta. Se ha documentado que este fenómeno se agudiza cuando se trata de búsquedas en áreas de estudio o temas altamente especializados o en campos relacionados con las ciencias sociales, donde la contextualización es primordial (Bender, *et al.*, 2021; Ji, *et al.*, 2023; Perkins y Roe, 2023; Mostafapour, *et al.*, 2024).

Por último, se plantean preocupaciones sobre la equidad y el acceso, tanto en las condiciones desiguales para desarrollo de la literacidad digital, como en todas

aquellas relacionadas con la accesibilidad, puesto que estas herramientas son accesibles para pocas economías, como las que se concentran en el Norte Global. En consecuencia, investigadores y académicos de muchos países del Sur Global quedan fuera del alcance de estas tecnologías, de la posibilidad de consultar algunas bases de datos bibliográficas y de tener su propia producción y difusión del conocimiento académico, generando una tendencia a mayor exclusión de ciertas áreas o grupos de investigación y acrecentando la brecha digital (Campolo, *et al.*, 2017; Bender, *et al.*, 2021; Perkins y Roe, 2023; Van Noorden y Perkel, 2023; UNESCO, 2024).

Perspectivas futuras y mejoras potenciales

Las perspectivas sobre el uso de asistentes de IAGen para la recuperación son prometedoras. Ya que esta tecnología ha demostrado un enorme potencial, fortalecen las habilidades de investigación, fomentan la colaboración, simplifican el proceso de búsqueda, pueden mejorar la accesibilidad y democratización del conocimiento y ayudan a identificar y combatir la desinformación científica. Se pretende que estos asistentes de IAGen se encuentren en un ciclo de mejora, tanto para la búsqueda precisa de términos como para la eficiencia de las revisiones sistemáticas y la evaluación constante con la finalidad de identificar y reducir el riesgo de sesgos (Mostafapour, *et al.*, 2024; Perkins y Roe, 2023; 2024).

No obstante, aunque los asistentes de IAGen proporcionan un impulso inicial, la lectura crítica y la selección final siguen siendo responsabilidad de los académicos. No se debe esperar que la herramienta haga todo el trabajo, sino que, por el contrario, hay que aplicar criterios propios y objetivos de investigación bien definidos para evaluar y seleccionar la literatura que ayude a resolver la necesidad de información; las herramientas de búsqueda de IA son aliadas valiosas en la investigación, pero no reemplazan la labor intelectual y el juicio humano. Al combinar la eficiencia de la tecnología con la capacidad de discernimiento se puede avanzar en el conocimiento y contribuir al progreso científico.

Las consideraciones éticas asociadas con el uso de estas herramientas son, hasta el momento, uno de los temas más relevantes y complicados de resolver pues se encuentra en plena discusión y desarrollo de políticas, normativas, regulaciones en todo el mundo; hasta la fecha sólo la Unión Europea ha publicado formalmente una ley para la regulación en el uso de la IA, la cual entró en vigor el 12 de julio de 2024 (European Union, 2024).

La mayoría de las discusiones de la ética se centran en el diseño y desarrollo, enfocándose en las normativas que debe tener la inteligencia artificial, pero pocas veces se habla de cómo la IA puede contribuir a la ética (Boucher, 2020), por ejemplo, los asistentes para la búsqueda de literatura podrían ayudar a mitigar el sesgo de información al usar algoritmos objetivos; estos sistemas pueden presentar una gama más amplia de perspectivas y fuentes, reduciendo la influencia de factores subjetivos en la selección de información (Moya y Eaton, 2023).

Otro tema ético importante es la transparencia y la equidad en el acceso a la información. Aunque los asistentes pueden proporcionar recomendaciones personalizadas y filtrar vastas cantidades de literatura, existe el riesgo de que la información presentada esté sesgada o limitada por los algoritmos subyacentes. Los desarrolladores deben garantizar que los criterios de selección sean claros y objetivos, evitando la promoción involuntaria de determinadas perspectivas o exclusiones injustas (Van Noorden y Perkel, 2023).

En el futuro inmediato, se requiere garantizar la privacidad y la seguridad de los datos. Estos asistentes pueden requerir acceso a información personal y datos de usuario para mejorar su precisión y personalización. Es crucial que se implementen medidas estrictas de protección de datos para evitar la exposición indebida de información confidencial y proteger la privacidad de los usuarios.

Otro aspecto importante es la responsabilidad en el uso de la información obtenida a través de estos asistentes. Los académicos deben ser conscientes de la necesidad de citar correctamente las fuentes y atribuir el mérito adecuadamente, especialmente cuando

la automatización de la búsqueda puede llevar a una menor supervisión humana en la recopilación de datos (Fernández-Samos-Gutiérrez, 2023).

Conclusiones

Es innegable que, ante el desarrollo del campo de la IAGen y el surgimiento de los asistentes para la recuperación de literatura científica, la forma en que accedemos y utilizamos la información académica se está transformando radicalmente, simplificando el proceso de búsqueda y la recopilación de información al usar como fuentes de información bases de datos científicas que organizan literatura con revisión por pares, verificada y validada; devolviendo resultados precisos y relevantes de manera rápida y eficiente, identificando tendencias emergentes en la investigación académica, analizando grandes volúmenes de literatura para detectar patrones y proporcionando a los usuarios una visión actualizada del panorama investigativo en su campo, personalizando recomendaciones.

No obstante, nos encontramos ante diversos retos que afrontar en el ámbito de la literacidad digital ya que, más allá de desarrollar habilidades y competencias, se requiere fomentar el pensamiento crítico, fortalecer actitudes y capacidades de los individuos para utilizar de manera ética la información para construir nuevos conocimientos y socializarlos a través de acciones constructivas y reflexivas.

Particularmente, los aspectos éticos implican un monitoreo constante y la suma de diversos actores educativos, gubernamentales, entre otros, para garantizar la transparencia en los algoritmos, la protección de la privacidad de los datos, la responsabilidad en el manejo de la información y la discusión sobre el posible impacto a largo plazo en la autonomía intelectual. La dependencia excesiva de estos asistentes podría potencialmente reducir las habilidades de búsqueda y evaluación objetiva de la información por parte de los usuarios, lo que podría limitar la diversidad de perspectivas y sesgar la investigación académica. ■

REFERENCIAS BIBLIOGRÁFICAS

- Aalto University Learning Centre, (2023). *Guide to information retrieval*. [En línea] <https://libguides.aalto.fi/informationretrieval>
- Abio, G. (2017). Formación digital de profesores: Una revisión del tema con énfasis en los modelos de competencias y literacidades digitales. *Caracol*, 13, 20–55.
- Barsha, S. and Munshi, S.A. (2024). Implementing artificial intelligence in library services: a review of current prospects and challenges of developing countries, *Library Hi Tech News*, 41(1), 7–10. <https://doi.org/10.1108/LHTN-07-2023-0126>
- Bender, E. M., McMillan-Major, A., Gebru, T, y Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA. <https://doi.org/10.1145/3442188.3445922>
- Boucher, P. (2020). *Artificial Intelligence: How does it work, why does it matter, and what can we do about it?*: Study. Brussels: European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU\(2020\)641547_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf)
- Boughanem, M., Akermi, I., Pasi, G., Abdulahhad, K. (2020). Information Retrieval and Artificial Intelligence. In: Marquis, P., Papini, O., Prade, H. (eds) *A Guided Tour of Artificial Intelligence Research*. Springer, Cham. https://doi.org/10.1007/978-3-030-06170-8_5
- Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI Now 2017 Report*. AI Now Institute, New York University. <https://ainowinstitute.org/>
- Carbonell-García, Carmen Elena, Burgos-Goicochea, Saby, Calderón-de-los-Ríos, Davis Osvaldo, & Paredes-Fernández, Oster Waldimer. (2023). La Inteligencia Artificial en el contexto de la formación educativa. Episteme Koinonía. *Revista Electrónica de Ciencias de la Educación, Humanidades, Artes y Bellas Artes*, 6(12), 152–166. Epub 18 de agosto de 2023. <https://doi.org/10.35381/e.k.v6i12.2547>
- Egger, A. E. y Carpi, A. (2009). El Uso de la Literatura Científica. *Visionlearning* Vol. POS-2 (7). <https://www.visionlearning.com/es/library/Proceso-de-la-Ciencia/49/El-Uso-de-la-Literatura-Cient%C3%ADfica/173>
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 20 September 2024 on harmonised rules for artificial intelligence and amending certain Union legislative acts. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689
- Fernández-Samos-Gutiérrez, Rafael. (2023). La inteligencia artificial en la redacción y autoría de publicaciones científicas. *Angiología*, 75(5), 281–283. <https://dx.doi.org/10.20960/angiologia.00512>
- Grewal, A., Kataria, H., & Dhawan, I. (2016). Literature search for research planning and identification of research problem. *Indian journal of anaesthesia*, 60(9), 635–639. <https://doi.org/10.4103/0019-5049.190618>
- Guo, W, y Caliskan, A. (2021). Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*, May 19–21, 2021, Virtual Event, USA. ACM, New York, NY, USA, 12. <https://doi.org/10.1145/3461702.3462536>
- Jardón Gallegos, M; Granizo Malusin, J; Yaselga Auz, W y Cocha Telenchana, M. (2024). Impacto de los asistentes virtuales de inteligencia artificial en el rendimiento académico de estudiantes universitarios. *Revista Social Fronteriza*; 4(4): e338. [https://doi.org/10.59814/resofro.2024.4\(4\)338](https://doi.org/10.59814/resofro.2024.4(4)338)
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A. y Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12). <https://doi.org/10.1145/3571730>
- Jones, R. H., y Hafner, C. A. (2021). *Understanding digital literacies* (2nd ed.). Routledge.
- Maldonado, C.E. (2020). *Teoría de la información y complejidad: La Tercera revolución científica*. Colección Complejidad y Salud, Vol. 4. Bogotá: Editorial Universidad El Bosque.
- Martínez Morales, Manuel. (2024). *Inteligencia Artificial... ¿Qué es eso?*. [En prensa]. Universidad Veracruzana, Dirección de Comunicación de la Ciencia. <https://www.uv.mx/cienciauv/blog/ia-quees/>

- Méndez-Ochaita, M. F. (2022). Multiliteracidades En La Era Digital: Conceptos Clave Desde Los Estudios De Literacidad Y Multimodalidad. *Revista Electrónica Leer, Escribir y Descubrir*, 1(11). <https://digitalcommons.fiu.edu/led/vol1/iss11/2>
- Michán, L., Álvarez, E., & Guillén, J. (2013). Introducción a la recuperación de literatura digital en ciencias biológicas. <http://www.ciib.unam.mx/libros/irldcb/index.html%5Cnhttp://repositorio.fcien-cias.unam.mx:8080/xmlui/handle/11154/103780>
- Moncada-Hernández, S. G. (2014). Cómo realizar una búsqueda de información eficiente. Foco en estudiantes, profesores e investigadores en el área educativa. *Investigación En Educación Médica*, 3(10), 106–115. [https://doi.org/10.1016/S2007-5057\(14\)72734-6](https://doi.org/10.1016/S2007-5057(14)72734-6)
- Mostafapour, M., Fortier, J. H., Pacheco K., Murray, H. y Garber, G. (2024). Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study, *JMIR AI*, 3: e56537. <https://doi.org/doi:10.2196/56537>
- Moya, B., & Eaton, S. E. (2023). Examinando Recomendaciones para el Uso de la Inteligencia Artificial Generativa con Integridad desde una Lente de Enseñanza y Aprendizaje. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 29(2). <https://doi.org/10.30827/relieve.v29i2.29295>
- Novelli, C., Casolari, F., Hacker, P., Spedicato, G., y Floridi, L. (2024). Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity. *Computer Law & Security Review*, 55, 106066. <https://doi.org/10.1016/j.clsr.2024.106066>
- Patel, A., Kusuma, L., Tantradi, S., Bekinal, S. & Ropashree, S. (2024). Breaking Language Barriers: A Global Translation Initiative. *Indiana Journal of Multidisciplinary Research*, 4(3), 16–23. <https://doi.org/10.5281/zenodo.12666921>
- Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2). <https://doi.org/10.53761/1.20.02.07>
- Perkins, M. y Roe, J. (2023). Decoding Academic Integrity Policies: A Corpus Linguistics Investigation of AI and Other Technological Threats. *Higher Education Policy*, 37:633–653. <https://doi.org/10.1057/s41307-023-00323-2>
- Perkins, M. y Roe, J. (2024). Academic publisher guidelines on AI usage: A ChatGPT supported thematic analysis. *F1000Research*, 12:1398. <https://doi.org/10.12688/f1000research.142411.2>
- Rzhetsky, A., Seringhaus, M., & Gerstein, M. (2008). Seeing a new biology through text mining. *Cell*, 134(1), 9–13. <https://doi.org/10.1016/j.cell.2008.06.029>
- Salvador Oliván, José Antonio y Arquero Avilés, Rosario. (2006). Una aproximación al concepto de recuperación de información en el marco de la ciencia de la documentación. *Investigación bibliotecológica*, 20(41), 13–43. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2006000200002&lng=es&tIng=es.
- Statista (2024). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. <https://www.statista.com/statistics/871513/worldwide-data-created/>
- Trujano Ruiz, P., Tovilla Quesada, V. y Dorantes Segura, J. (2008). Educación digital. Los retos del nuevo siglo. *Revista Electrónica de Psicología Iztacala*, 11(3). <http://revistas.unam.mx/index.php/repi/article/viewFile/18560/17618>
- UNESCO. (2024). *Guía para el uso de ia generativa en educación e investigación*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. <https://creativecommons.org/licenses/by-sa/3.0/igo/deed.es>
- Van Noorden, R. y Perkel, M. (2023). AI And Science: What 1,600 Researchers think. *Nature*, 621:672–675. <https://doi.org/10.1038/d41586-023-02980-0>

ANEXO 1: Asistentes de IAGen para la investigación académica

Herramienta	Descripción	Base de datos	Otras características	Precio
Consensus	Se basa en modelos de LLMs y búsqueda vectorial para mostrar resultados relevantes, sintetiza información tanto por temas como por artículo. Los resultados están basados en artículos reales.	Semantic Scholar	<ul style="list-style-type: none"> Resúmenes de artículos generados por GPT4. Medidores de consenso para preguntas de sí/no. Resúmenes de estudios (métodos, resultados, etcétera). Copilot de Consensus (responde preguntas, redactar contenido, entre otras funciones). Complemento de Zotero. 	Planes gratuitos y de pago
Elicit	Utiliza LLMs para extraer datos y resumir artículos de investigación. También permite seleccionar artículos relevantes, buscar similares y extraer detalles de los artículos en una tabla organizada.	Semantic Scholar	<ul style="list-style-type: none"> Permite subir archivos PDF (<i>solo material sin derechos de autor</i>) y obtener resúmenes. Lista de conceptos (para buscar métodos, conjuntos de datos y el mismo concepto en diferentes temas). Complemento de Zotero. 	Planes gratuitos y de pago
Keenious	Recomienda artículos de investigación relevantes analizando el documento.	OpenAlex	<ul style="list-style-type: none"> Resalta el texto en un archivo para buscar el concepto y/o tema específico. 	Planes gratuitos y de pago
ORKGAsk	Utiliza búsquedas vectoriales, LLMs y gráficos de conocimiento.	CORE Almacenamiento de Vectores: Qdrant. Modelo de Incorporación: Nomic Modelo de Lenguaje: LLM Mistral Instruct 7B v0.2	<ul style="list-style-type: none"> Presenta los resultados en una tabla para brindar a los usuarios una visión general de: preguntas de investigación del usuario, palabras clave, método, resultados y conclusiones. Exporta resultados (CSV y LaTeX) 	De acceso abierto
PerplexityAI	Plantea directamente preguntas y recibe respuestas precisas. No está diseñado específicamente para la investigación académica; sin embargo, los usuarios pueden elegir el enfoque "Académico".	ChatGPT ClaudeAI Meta LLaMA	<ul style="list-style-type: none"> Copilot (solo para usuarios registrados): hace preguntas sobre las consultas para adaptar mejor los resultados a las necesidades de los usuarios. 	Planes gratuitos y de pago
SciSpace	Encuentra, comprende y conecta literatura científica: <i>"Por cada artículo que leas, obtén explicaciones simples y respuestas de ia y descubre una red de artículos conectados y relevantes".</i>	Semantic Scholar Directorio de Libros de Acceso Abierto Directorio de Revistas de Acceso Abierto	<ul style="list-style-type: none"> Ofrece una plataforma multipropósito para la investigación y publicación académica. Análisis de artículos. Verificación de plagio. 	Planes gratuitos y de pago
Semantic Scholar	<i>"Busca, encuentra y comprende la literatura científica más relevante para tu trabajo."</i>	Semantic Scholar	<ul style="list-style-type: none"> Biblioteca personal. Alertas para nuevas publicaciones. Resúmenes generados por ia (actualmente solo para ciencias de la computación y biomedicina). 	De acceso libre
Undermind	Utiliza un complejo algoritmo de búsqueda de cuatro pasos para <i>"imitar el cuidadoso y sistemático proceso de búsqueda de un humano"</i> .		<ul style="list-style-type: none"> Resúmenes del tema basados en los artículos que se encontraron. Refinamiento de los resultados basados en temas específicos. 	Planes gratuitos y de pago
Citation Gecko	<i>"Busca en la red de citas artículos conectados, permitiéndote identificar rápidamente artículos importantes que quizás hayas pasado por alto."</i>	Indefinido	<ul style="list-style-type: none"> Importa desde Bibtex, Zotero y Mendeley. Diseño de línea de tiempo frente a diseño de red. 	De acceso libre

Herramienta	Descripción	Base de datos	Otras características	Precio
Connected Papers	"Encuentra y explora artículos relevantes para tu trabajo". Proporciona un mapa visual de otros artículos relacionados (temática o bibliográficamente).	Semantic Scholar Paper Corpus (ODC-BY).	<ul style="list-style-type: none"> • Visualiza los resultados como un gráfico o como una lista. • Conteo de citas disponible. 	Planes gratuitos y de pago
Inciteful	Explora la literatura académica utilizando análisis de redes de citas.	Inciteful.xyz	<ul style="list-style-type: none"> • Conector de Literatura: los usuarios pueden explorar cómo se conectan dos cuerpos de literatura entre sí ingresando dos artículos. • La herramienta muestra los caminos más cortos entre los artículos y permite la interacción con el gráfico. • Busca y filtra artículos. • Complemento de Zotero. 	De acceso libre
Litmaps	Recomienda artículos basados en un artículo inicial de partida. Los usuarios pueden buscar un tema, seleccionar un artículo y Litmaps encontrará artículos relacionados relevantes.	Crossref Semantic Scholar OpenAlex	<ul style="list-style-type: none"> • Organiza artículos en colecciones. • Encuentra más artículos relacionados con "Búsqueda de Descubrimiento". • Guarda búsquedas, exportas y comparte hallazgos. • Conteo de citas disponible. 	Planes gratuitos y de pago
Local Citation Network	Realiza una revisión de literatura utilizando metadatos de OpenAlex (OA), Semantic Scholar (S2), OpenCitations (OC) y Crossref (CR). Visualiza subconjuntos de la red de citas global y se sugieren las referencias más citadas para su revisión.	OpenAlex Semantic Scholar OpenCitations CrossRef	<ul style="list-style-type: none"> • Descargas de archivos en formato RIS, CVS y JSON. • Muestra resultados tanto como red gráfica y tabla. 	De acceso libre
Research Rabbit	Está basada en el mapeo de literatura basada en citas.	Indefinido	<ul style="list-style-type: none"> • Alertas personalizadas. • Crea colecciones personalizadas. • Descubre redes de autores. • Complemento de Zotero. • Exportar artículos en formato BibTex, RIS, CSV. • Conteo de citas. 	De acceso libre
Open Knowledge Maps	Proporciona una visión general de un tema al mostrar las principales áreas temáticas y los documentos relacionados con cada área.	PubMed, OpenAIRE.	<ul style="list-style-type: none"> • Proporciona texto completo en PDF para artículos de acceso abierto. • Exporta en formato BibTeX. 	De acceso libre
Scite	Especializada en "cómo los artículos fueron tratados en otros artículos", identifica lagunas de investigación y controversias dentro de un campo.	Unpaywall, PubMed Fatcat Repositorios universitarios Revistas de acceso abierto	<ul style="list-style-type: none"> • Complemento de Zotero. 	Planes gratuitos y de pago
VOSViewer	Diseñada para construir y visualizar redes bibliométricas, basadas en citas, acoplamiento bibliográfico, co-cita o relaciones de coautoría. También "ofrece la funcionalidad de extracción de texto que puede usarse para construir y visualizar redes de coocurrencia de términos importantes extraídos de un cuerpo de literatura científica."	Web of Science Scopus Dimensions Lens PubMed Crossref Europe PMC OpenAlex Semantic Scholar OpenCitations WikiData	<ul style="list-style-type: none"> • Cuenta con una versión de escritorio y una versión web. 	De acceso libre

Tomada y modificada de: Artificial Intelligence for Academic Research (2024). University Library Eberhard Karls Universität Tübingen. <https://uni-tuebingen.de/en/facilities/university-library/learning-working/workshops-consultations-and-guided-tours/literature-research-with-ai/#c2071131>