

# Los *chatbots* de IA generativa como fuentes de tráfico web: aproximación desde la analítica web de un portal de revistas

## Generative AI chatbots as sources of web traffic: web analytics insights from a journal portal

Cristóbal Urbano

Urbano, Cristóbal (2024). "Los *chatbots* de IA generativa como fuentes de tráfico web: aproximación desde la analítica web de un portal de revistas". *Anuario ThinkEPI*, v. 18, e18e31.

<https://doi.org/10.3145/thinkepi.2024.e18a31>

Publicado en *IweTel* el 22 de noviembre de 2024

### Cristóbal Urbano

<https://www.directorioexit.info/ficha55>

<https://orcid.org/0000-0003-0935-6436>

Universitat de Barcelona

Departament de Biblioteconomia, Documentació i

Comunicació Audiovisual

[urbano@ub.edu](mailto:urbano@ub.edu)



**Resumen:** Reflexión sobre el papel de los *chatbots* de IA generativa en el tráfico web referido que llega a las revistas académicas. Se establecen escenarios sobre la influencia que pueden tener en el comportamiento informacional de búsqueda, dentro del entorno académico, esas nuevas herramientas basadas en *Large Language Models* (LLM) y en la generación de respuestas directas resultantes de un proceso iterativo de conversación con un *chatbot* de IA. Mediante la exploración de datos de analítica web obtenidos con *Google Analytics* para el portal *RACO*, se realiza una primera aproximación al bajo volumen de visitas que, hasta la fecha, parece que se están recibiendo en revistas académicas procedentes de *chatbots* de IA. Se formulan hipótesis que pueden explicar ese bajo volumen de tránsito, dando por descontado que el uso de dichas herramientas de IA es importante y creciente en el mundo académico. Por último, se plantea cuál ha de ser el papel del *Generative Engine Optimización* (GEO) en tanto que estrategia de posicionamiento de contenidos académicos en este nuevo tipo de herramientas de búsqueda alimentadas con IA generativa y, en qué medida han de ser un complemento oportuno a las acciones ya establecidas de *Academic Search Engine Optimization* (ASEO). Se invita al debate mediante una serie de preguntas abiertas sobre cómo las nuevas herramientas de IA pueden estar cambiando el panorama de la búsqueda y uso de información bibliográfica, y cómo eso puede afectar a las revistas académicas en sus estrategias visibilidad y posicionamiento.

**Palabras clave:** Inteligencia artificial; *Chatbots* de IA generativa; Comportamiento informacional; Búsqueda bibliográfica; Revistas académicas; Analítica web; *Google Analytics*; SEO; GEO; ASEO.

**Abstract:** This paper is a reflection on the role of generative AI chatbots in the referred web traffic that reaches academic journals. Scenarios on informational search behaviour within the academic environment are explored, regarding the influence of these new tools based on Large Language Models (LLM) and on the generation of direct responses resulting from an iterative process of conversation with an AI chatbot.

By exploring web analytics data obtained with Google Analytics for the RACO portal, a first approach is presented to the low volume of visits that, to date, it seems that academic journals are receiving from AI chatbots. Hypotheses are formulated to explain this low volume of traffic, assuming that the use of these AI tools is important and growing in the academic world. Finally, the role of Generative Engine Optimization (GEO) as a strategy for academic content optimization in generative AI-powered search tools is discussed, to decide whether it should be a suitable complement to the already established Academic Search Engine Optimization (ASEO) actions. Discussion is invited through a series of open questions on how new AI tools may be changing the landscape of bibliographic information search and use, and how that may affect scholarly journals in their visibility and optimization strategies for traffic acquisition.

**Keywords:** Artificial Intelligence; Generative AI chatbots; Information behaviour; Bibliographic search; Academic journals; Web analytics; *Google analytics*; SEO; GEO; ASEO.

Hay pocas dudas de que con la apertura al público de *ChatGPT*, en la todavía cercana fecha del 30 de noviembre de 2022, se produjo una gran transformación en los hábitos de consumo de información de colectivos bien diversos, que en pocos meses tuvieron ante sí esa y muchas otras herramientas de IA generativa (entre otras, *Gemini*, *Copilot*, *Claude*, *Perplexity*, *You.com*, *Consensus*, etc.). Se trata de un proceso que percibimos como muy presente en las conversaciones informales en el entorno académico, así como en infinidad de debates más formales sobre cómo la IA incide en la forma de aprender, de enseñar y de investigar; en último extremo, en la forma de leer, pensar, escribir o crear.

Más allá de las percepciones, contamos con datos estadísticos de captación de usuarios que reportan las empresas productoras, así como estudios de audiencia realizados mediante encuestas y métodos indirectos de análisis de la popularidad de apps y sitios web, que ayudan a describir la “ola” de la tendencia en la que estamos inmersos. Por otra parte, también existe un buen número de artículos de descripción de experiencias y situaciones en el ámbito académico, referidos sobre todo a *ChatGPT* (**Goyanes; Lopezosa, 2024; Lopezosa; Goyanes, 2024**). Por último, son ya numerosos y relevantes los documentos de toma de posición institucional y de recomendaciones que emanan de importantes organismos, como por ejemplo Unesco (**Miao; Holmes, 2023**) o la CRUE (**Cruz-Arquero et al., 2024**), entre otros muchos. En resumen, todo ello, junto a la abundante atención que la prensa le ha prestado al tema, apunta a una gran extensión en el uso de las herramientas de IA, con importante impacto en la vida académica. Sin embargo, todavía no hay revisiones sistemáticas que consoliden los pocos estudios publicados sobre comportamiento informacional de búsqueda en el ámbito académico (p. ej. **Deschenes; McMahon, 2024**), que muestren cómo, por qué y con qué consecuencias se están usando estas nuevas herramientas.

En ese marco, la presente nota *ThinkEPI* pretende estimular el debate sobre cómo las nuevas herramientas de IA pueden estar cambiando el panorama de la búsqueda y uso de información, y cómo eso puede afectar a las revistas académicas de acceso abierto en sus estrategias de visibilidad y posicionamiento, tanto mediante acciones de SEO orientadas a buscadores generales, como de ASEO (*Academic Search Engine Optimization*) en buscadores académicos o de indexación en bases de datos bibliográficas.

## Comportamiento informacional y herramientas que ofrecen respuestas directas

Dentro del abanico de herramientas de IA tipo *chatbot*, y haciendo una aproximación muy genérica, podemos distinguir tres grados de respuesta directa en relación con la inclusión, o no, de enlaces o de referencias bibliográficas a documentos.

En primer lugar, aquellas herramientas que por defecto no ofrecen las fuentes en las que sustentan sus respuestas y que en caso de que les solicitemos “bibliografía” para ampliar lo que nos dicen, pueden “alucinar” creando referencias bibliográficas totalmente verosímiles pero inventadas, como por ejemplo *Claude.ai* o *ChatGPT4o-mini* (en su versión gratuita sin conexión a *ChatGPT-Search*).

En segundo lugar, aquellas que estando también orientadas a la generación de respuestas ofrecen por defecto enlaces a fuentes en las que se apoyan, p. ej. *Copilot* o *ChatGPT4o* (cuando se conecta a *ChatGPT-Search* en versiones de pago).

En tercer lugar, podríamos considerar aquellas herramientas que se plantean como motores de búsqueda de nueva generación, que a partir de un *prompt* generan estrategias para recuperar contenido relevante que sirve para destilar respuestas en las que se citan las fuentes, p. ej. *You.com*, *Perplexity*, *Consensus*. Una variación destacable de esta tercera tipología serían los desarrollos de búsqueda conversacional asistida por IA sobre el contenido de bases de datos bibliográficas tradicionales, p. ej. *Scopus IA*. Bajo esta variante también podríamos incluir algunas herramientas como *Undermine*, *ScholarAI* o *Elicit*, que se ofrecen como asistentes de búsqueda bibliográfica y condensación de los contenidos recuperados, en tanto que soluciones para la realización de revisiones bibliográficas exploratorias, de cara a la redacción inicial de estados de la cuestión.

En este nuevo escenario, resulta pertinente preguntarnos en qué medida se puede estar transformado el comportamiento informacional de búsqueda bibliográfica, elemento central de la cultura académica, entendido en el sentido de lo que **Gorichanaz** (2024) denomina el “desarrollo de la virtud intelectual” en relación con las búsquedas en línea. Hablamos de nueva transformación, ya que los hábitos de búsqueda ya experimentaron un profundo cambio con el uso abrumador de *Google* y de *Google Scholar* y el declive de las bases de datos de indización y resumen. En todo caso, tanto en el paradigma pre-*Google* como en el actual, la búsqueda se venía contemplando como un proceso de satisfacción de las necesidades de información a través de la lectura “humana” de documentos seleccionados a partir de referencias (enlaces, si hablamos del entorno web); referencias que se obtenían por la vía de bibliografías, catálogos, bases de datos, buscadores web, recomendaciones de lecturas o seguimiento de la bibliografía citada.

Pues bien, el punto nuclear que estaría marcando el cambio de tendencia en el comportamiento informacional en el entorno de la IA generativa es que las nuevas herramientas ofrecen al usuario un modelo de satisfacción de sus necesidades de información que podríamos denominar “tipo oráculo” (**Messeri; Crockett, 2024; Via, 2023**) basado en respuestas directas sintéticas y no tanto en los enlaces a documentos. Se trata de un fenómeno que ya se venía estudiando con anterioridad, fruto de la aparición en buscadores “tradicionales” como *Google* de los *snippets* enriquecidos que apuntan con información directa y de forma preferente hacia una fuente, o que ofrecen también una lista de preguntas tipo, con respuestas en forma de extractos de sitios web, incrustadas en la *SERP* del buscador (**Potthast; Hagen; Stein, 2020; Strzelecki; Rutecka, 2020**). En ambas situaciones, los usuarios pueden juzgar innecesario visitar los enlaces facilitados. Por este efecto de las respuestas directas, los expertos consultados en el estudio de *Ithaca S+R* sobre impacto de la IA en el ecosistema editorial pronostican un descenso de las visitas a las revistas (**Bergstrom; Ruediger, 2024**).

En ese mismo sentido, en una entrevista reciente en el diario *Avui* con Álex Hinojo, reconocido wikipedista y miembro de la dirección de *Amical Wikipedia* entre 2012 y 2018, se planteaba que las nuevas herramientas de búsqueda basadas en IA generativa supondrán una reducción del tráfico en la mismísima *Wikipedia*:

“Ya hace tiempo que la tecnología responde preguntas sencillas sin ir a *Wikipedia*. Ahora, con buscadores basados en modelos de lenguaje, ni siquiera es necesario ir a la fuente. Le pides:

¿qué ha pasado hoy? O: Explícame la Revolución Francesa. Y lo hace en tono divulgativo. Es muy goloso. La *Wikipedia* depende de voluntarios; si bajan los lectores, bajarán los editores. [...]. La *Wikipedia* pasará de ser fuente diaria de consulta, a un recurso de verificación en una internet dominada por la automatización y la IA” (**Gordillo-Bernárdez**, 2024).

## Hacia un nuevo tipo de buscador web: los Generative Engines

Ciertamente, a la vista de cómo van evolucionando los diversos *chatbots* de IA, no podemos decir que los enlaces estén totalmente ausentes en las respuestas, pues cada vez son más los que ofrecen búsquedas en la Web en general, o en conjuntos de información académica, que se sustentan mediante el mismo proceso conversacional entre el usuario y la aplicación, pero que muestran enlaces integrados en algún tipo de narrativa producto de una recuperación generativa de información (**Li et al.**, 2024). En este sentido, todo indica que la aparición de *ChatGPT-Search* puede marcar un antes y un después en las búsquedas en la Web (**Nafría; Nafría**, 2024).

<https://openai.com/index/introducing-chatgpt-search>

En primer lugar, por la amenaza que representa para los buscadores tradicionales, y en especial para el monopolio de facto que *Google* ostenta en numerosos países, como por ejemplo España; en segundo lugar, por el efecto en el descenso de visitas a determinados sitios web producto del modo de seleccionar y presentar los resultados que tienen herramientas de búsqueda como *ChatGPT-Search*.

Se habla ya de un nuevo tipo de buscador al que se denomina *Generative Engine* y que usa modelos de IA generativa para reunir y resumir información con la que responder las consultas de los usuarios (**Aggarwal et al.**, 2024). *Google* también está avanzando en esa línea con desarrollos experimentales, no del todo exitosos todavía, como *Google SGE* (**Smith**, 2024; **Withorn**, 2023) que pretenden transformar el proceso de búsqueda hacia el tipo de experiencia de usuario que ofrecen *chatbots* tipo como *ChatGPT*. Semejantes movimientos anticipan un cambio en el mercado de las búsquedas que afectará a los creadores de contenido, entre los que se cuentan los editores de revistas. Unos creadores de contenido que se debatirán entre posicionar sus sitios web en las nuevas plataformas alimentadas por IA y la pérdida de visitas que puede representar que dichas plataformas ofrezcan directamente la respuesta a las necesidades de información, sin necesidad de visitar el sitio web original.

Es por ello que las tendencias observadas entre usuarios y proveedores de plataformas de IA, harán que los creadores de contenidos se planteen estrategias de lo que se ha venido en denominar *Generative Engine Optimisation* (GEO) (**Aggarwal et al.**, 2024) como complemento a sus acciones de SEO. En cualquier caso, parece que todo indica que se tendrán que enfrentar a un casi seguro menor número de visitas efectivas a sus sitios web. Como se dice en una pieza reciente de prensa en *The Guardian*:

“Formule una pregunta a un motor de búsqueda y éste le devolverá una larga lista de páginas web. La mayoría de los usuarios eligen las primeras, pero incluso las últimas obtienen tráfico. Los *chatbots*, en cambio, sólo mencionan los cuatro o cinco sitios web de los que obtienen la información como referencia” (**Bains**, 2024).

En efecto, en todos los casos las respuestas de los *chatbots* de IA tienden a ser mucho más selectivas en cuanto al número de enlaces que ofrecen, lo que unido a su capacidad de respuesta directa puede inhibir la visita a los enlaces propuestos. El resultado de todo ello ya se está viendo como una amenaza para la captación de tránsito en el canal “organic search”, según la terminología al uso en analítica web. No es extraño, pues, que en el mundo de la consultoría en posicionamiento web y marketing digital se esté hablando de la necesidad de hacer el tránsito desde el SEO (*Search*

Engine Optimization) al GEO (Generative Engine Optimization) (Bains, 2024), algo que de forma más cruda algunos autores denominan manipulación de LLMs para incrementar la visibilidad de productos (Kumar; Lakkaraju, 2024).

## Analítica web del tráfico de referencia en un portal de revistas: RACO

Partiendo de este contexto y de mi interés por la analítica web aplicada a revistas académicas (Vitela-Caraveo; Urbano, 2020), me ha parecido interesante hacer una primera exploración de la presencia de los chatbots de IA en el tráfico referido que llega al sitio web de un portal de revistas académicas. Gracias a la colaboración del Área de Aprendizaje, Investigación y Ciencia Abierta del Consorci de Serveis Universitaris de Catalunya (CSUC), he tenido acceso a los datos de la propiedad Google Analytics (GA) correspondiente a RACO: *Revistes Catalanes amb Accés Obert*.

<https://raco.cat>

Además, con la colaboración de la Facultat d'Informació i Mitjans Audiovisuals de la Universitat de Barcelona, he utilizado también los datos de la revista *BiD*, para así tener un punto de comparación entre un portal como RACO gestionado con OJS, y una revista concreta cuyos artículos están todos disponibles en html (excepto los correspondientes al año 2024).

<https://bid.ub.edu>

Con el comentario de estos datos se pretende aportar evidencias al debate sobre horizontes de futuro en el posicionamiento web y en la captación de audiencia por parte de revistas académicas. A partir de esta primera impresión retrospectiva, la idea sería ir viendo cómo evolucionan en los años venideros estas nuevas fuentes de tráfico en comparación con otras más consolidadas, tanto en RACO como en otros portales de revista interesados.

Se ha trabajado con los datos que cubren el período del 1/12/2022 al 17/11/2024: esto es, desde el día siguiente al lanzamiento de *ChatGPT* hasta el momento justo de redacción de esta nota. A largo de estos 23 meses y medio, según los datos de GA, se registraron en RACO un total de 3.981.591 sesiones, destacando el medio "Referral" como el que contaba con un mayor porcentaje de interacción (indicador relacionado con la calidad de las visitas) y el medio "Organic" como el que obtiene el mayor porcentaje (46,27%) sobre el total de visitas (tabla 1). Bajo la configuración de GA que tiene establecido RACO para el medio "Organic", aparecen los buscadores web generalistas. Entre ellos, se observa un dominio casi absoluto de *Google*, pues genera el 99,95% del total del tráfico que llega por dicho el medio. Otros buscadores como *Bing*, *Baidu*, *Ecosia*, *DuckDuckGo*, *Yahoo*, tienen una cuota insignificante.

Tabla 1. Métricas de las sesiones en RACO según medio de la sesión

Medio de la sesión	Sesiones	% sobre total	% interacciones	Duración media sesión
Organic	1.842.334	46,27	74,60	4 min y 15 s
Referral	1.315.722	33,05	81,65	5 min y 26 s
Direct	815.354	20,48	73,82	5 min y 02 s
Otros	8.181	0,21	n/a	n/a
Total	3.981.591	100,00	77,00	4 min y 43 s

El medio "Referral", con el 33,05% del tráfico, tiene un peso muy notable en RACO. En la configuración de GA para RACO, corresponde al tráfico que proviene de sitios web de muy diversa naturaleza: buscadores académicos como *Google Scholar*, bases de datos bibliográficas como *Dialnet*, herramientas de descubrimiento de bibliotecas, redes sociales como *Twitter* o *Facebook*, páginas de contenido

como *Wikipedia*, etc. Esto es, bajo este medio se contabiliza el tráfico que proviene de cualquier sitio web que enlace a *RACO* y que no esté filtrado por *GA* como un buscador asignado al canal “Organic”. Por ello incluye *ChatGPT*, *Perplexity* y cualquier otra herramienta tipo *chatbot* de IA generativa.

Para un portal de revistas es muy interesante poner la lupa sobre el tráfico que se recibe de sitios web clasificados como medio “Referral”, ya que es el que nos puede ayudar a ver el papel que tienen en la adquisición de visitas diversas fuentes relacionadas con lo que podríamos denominar búsquedas de información bibliográfica más claramente caracterizables como tal. Pese a que sabemos que también se utiliza un buscador generalista como *Google* a la hora de iniciar una exploración bibliográfica, las intenciones de búsqueda más claramente orientadas a la investigación bibliográfica se acaban dirigiendo a fuentes como *Google Scholar* y otras del ámbito más tradicional del control bibliográfico. Eso explica que bajo dicho medio se registre el mayor porcentaje de interacción y las sesiones duren más.

Dentro del mismo, se han filtrado las sesiones que tienen su origen en los *chatbots* de IA activos en *RACO* (tabla 2). Los dominios que se han identificado dentro de ese tipo de fuentes con tráfico en *RACO* han sido los siguientes: chat.mistral.ai, chatgpt.com, chat.openai.com, claude.ai, consensus.app, copilot.microsoft.com, gemini.google.com, gptgo.ai, Iris.ai, komo.ai, openread.academy, perplexity.ai, phind.com, poe.com, typeset.io, you.com.

Tabla 2. Número de sesiones procedentes del agregado de fuentes tipo *chatbot* IA para el período analizado en *RACO* y en *BiD*

Fuente	Sesiones en RACO	Sesiones en BiD
perplexity.ai	1.173	444
chatgpt.com + chat.openai.com	393	109
you.com	346	162
gemini.google.com	54	17
typeset.io [SciSpace]	45	11
claude.ai	39	1
poe.com	37	13
copilot.microsoft.com	23	60
consensus.app	12	0
komo.ai	9	3
chat.mistral.ai	5	0
gptgo.ai	4	2
iris.ai	1	0
openread.academy	1	0
phind.com	1	2
Total	2.143	824

En lo referente a la comparación con *BiD*, sería interesante investigar qué influencia puede tener en el posicionamiento de las revistas en las herramientas de IA, que el contenido de los artículos sea html o pdf, o que el gestor de contenidos (CMS) sea de un tipo u otro. Se ha de tener en cuenta que *RACO* contiene 585 revistas y registró en el período analizado casi 4 millones de sesiones, mientras que *BiD* es una única revista con un tráfico en el mismo período de 273.635 sesiones. Ciertamente, en proporción al tráfico total, todo indica que la captación de tráfico de dichas fuentes IA fue mayor en el caso de la revista *BiD* (tabla 2).

En ambos casos son cifras muy discretas. Especialmente irrelevantes en el conjunto de los casi 4 millones de sesiones que registra *RACO* en el período analizado. De todas maneras, las 2.143 sesiones generadas desde fuentes tipo *chatbot* IA se encuentran cerca de las 2.695 del buscador

Bing, y muy por encima de otros buscadores como *Ecosia* y *Duckduckgo*. En cuanto a la evolución a lo largo del tiempo en el caso de *RACO*, se observa una tendencia continuada al alza (Figura 1), en la que hemos observado que el mayor crecimiento en los últimos dos meses y medio está vinculado a *ChatGPT*, seguramente fruto de la activación de la función *ChatGPT-Search* por parte de los usuarios de pago.

Si se compara el volumen total del tráfico recibido desde chatbots de IA durante el período analizado con el de las principales fuentes del medio “Referral” (tabla 3), se ve con mayor crudeza el bajo nivel de tráfico procedente desde sitios como *ChatGPT* o similares, que suponemos se usan con gran intensidad por los diversos grupos del colectivo académico y estudiantil, pero que hasta el presente no aportan un tráfico comparable al de otras fuentes.

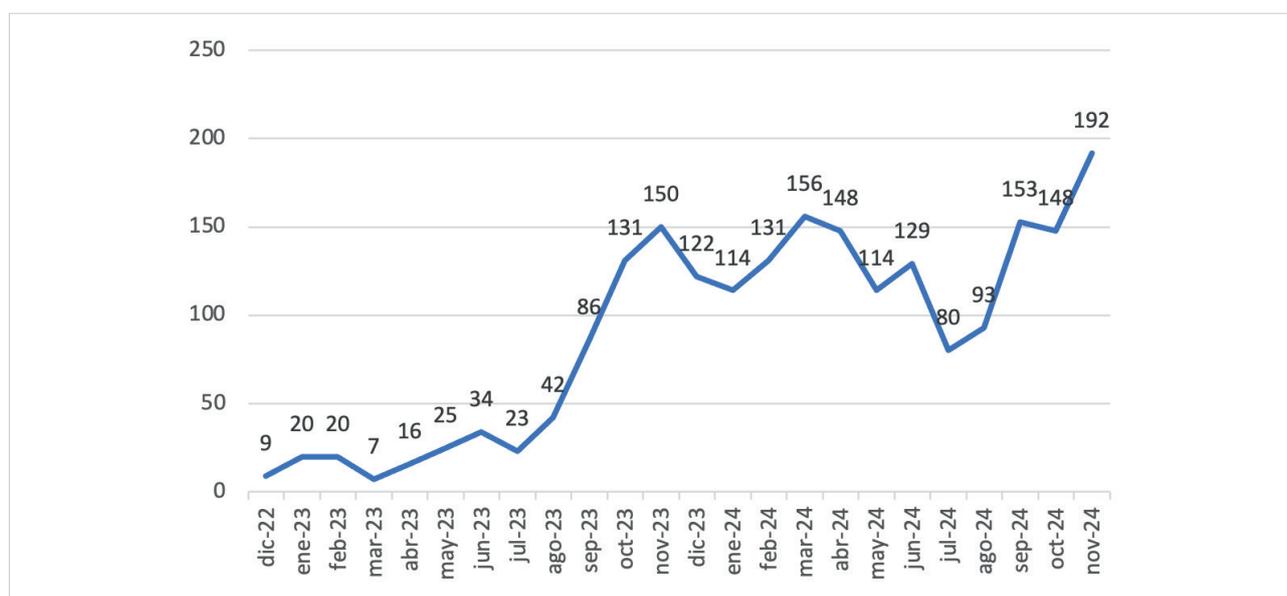


Figura 1. Evolución del número de sesiones en *RACO* procedentes del agregado de fuentes tipo chatbot IA

Tabla 3. Total de sesiones en *RACO* de diversas agrupaciones de fuentes del medio “Referral”.

Fuente de la sesión	Sesiones	% sobre el total del medio “Referral”
Google Scholar	650.478	49,44
Dialnet	242.156	18,40
Discovery tools de las bibliotecas de las universidades catalanas	54.357	4,13
Campus virtuales de las universidades catalanas	41.972	3,19
Wikipedia	18.632	1,42
Facebook	16.304	1,24
Twitter (X)	14.755	1,12
DOAJ	6.885	0,52
scimagojr.com	4.095	0,31
Selección de bases de datos de indización por citas (WoS, Scopus, Dimensions, OpenAlex, Lens, SemanticScholar, Scite)	3.543	0,27
Chatbots IA	2.143	0,16
Otros	260.402	19,79
Total medio “Referral”	1.315.722	100,00

Ciertamente, el volumen de tráfico que está llegando por las fuentes tipo *chatbot* IA es mínimo en comparación con la mayoría de otras fuentes del medio “Referral”. Es interesante observar que también representan una fracción muy pequeña frente al tráfico que llega de fuentes más o menos tradicionales del ámbito del control bibliográfico donde *RACO* está muy bien cubierta, como *Dialnet* o *DOAJ*, o las “*discovery tools*” de las bibliotecas universitarias catalanas. Sin embargo, sí que comienza a tener un valor relevante en su comparación con fuentes de control bibliográficas dirigidas a perfiles más iniciados en la investigación, como las bases de datos de indización por citas.

Asumiendo que el uso de herramientas tipo *ChatGPT* se estima creciente entre la población general, y de forma especial entre el público universitario, podríamos formular algunas hipótesis sobre los factores que podrían explicar el bajo volumen de visitas registradas en *RACO* procedentes de *chatbots* de IA:

- la ausencia de enlaces en un buen número de las respuestas obtenidas en dichas herramientas,
- una presencia muy selectiva de enlaces cuando sí aparecen,
- el efecto de las respuestas directas en el no seguimiento de enlaces por parte de los usuarios y, por último,
- la dificultad de posicionamiento de las revistas de *RACO* en los modelos de lenguaje y las fuentes de contenido que usan los *chatbots*.

Esperemos en el futuro poder discernir el peso de cada uno de esos factores, ya que esa información será útil en la toma de decisiones sobre acciones de optimización en la captación de visitas.

### **Para concluir y dar paso al debate**

A la vista de los datos, se podría afirmar que, desde el punto de vista del acceso y lectura de los artículos de revista en su sitio web original, los *chatbots* de IA no son actores emergentes de la búsqueda bibliográfica. Llevado al extremo, son justamente la expresión de lo que podríamos denominar búsqueda no-bibliográfica.

La explicación es sin duda multifactorial, pero si vamos a la raíz del comportamiento informacional en el que se fundamenta el uso de los *chatbots* de IA por parte de los usuarios, podemos suponer que se basan en el principio del “*good enough*” (Connaway; Dickey; Radford, 2011; Gorichanaz, 2024), según el cual, un usuario puede considerar que es conveniente dejar de buscar una vez se dispone de un resultado que juzgue como suficientemente bueno, en lugar de insistir en la búsqueda de la mejor respuesta posible. El modelo de interrogación tipo “oráculo” de los *chatbots* de IA encaja muy bien en el comportamiento basado en la aceptación de respuestas suficientemente buenas, especialmente cuando se asume que la IA, por su gran capacidad de proceso de contenidos, ya se habrá “leído” todo lo que fuera preciso. Muchos usuarios pueden dar por descontado una especie de “poderío lector” de estas herramientas, pero ciertamente no todas son iguales ni por la forma en la que han sido entrenadas, ni por los contenidos con los que se ha realizado el entrenamiento, ni por el acceso a fuentes de búsqueda cuando la herramienta contempla esa funcionalidad.

Ciertamente, las derivadas del debate sobre el uso de los *chatbots* de IA en la búsqueda de información por parte de usuarios del entorno universitario supera el propósito de esta nota, enfocada a la analítica web del tráfico captado por un portal de revistas y su utilidad para optimizar la captación de visitas. Pero para no obviar del todo otros aspectos, se podría hablar de asuntos bien importantes como: la formación de usuarios para un uso óptimo y ético de estas herramientas, la evaluación de la cobertura de contenidos y la calidad de las respuestas de cada uno de los muchos de productos diferentes que tenemos a nuestra disposición, las funcionalidades específicas que al-

gunas herramientas concretas tienen para la realización de revisiones bibliográficas o para el envío de alertas de novedades acordes con nuestros intereses, etc.

En lo referente al tema de esta nota, cabe finalizar diciendo que la creación de contenido original, riguroso, de calidad, arbitrado y seleccionado, ha sido hasta la fecha una característica de la comunicación científica de calidad. Un proceso que depende de la salud de un ecosistema editorial que utiliza una combinación de datos de audiencia y de impacto para evaluar su prestigio, su viabilidad y sus opciones de mejora. Para finalizar, les dejo algunas preguntas para estimular el debate al respecto:

- ¿Es posible y conveniente posicionar contenidos de una revista mediante técnicas de GEO en herramientas de IA? ¿Cómo se hace?
- ¿Se reducirán las visitas que reciben las revistas a causa de los cambios en los comportamientos de búsqueda en el entorno académico y de la evolución de las herramientas de búsqueda web hacia las respuestas directas y la limitación en el número de enlaces mostrados?
- ¿Se puede seguir optimizando la adquisición de tráfico en fuentes del medio “Referral” vinculadas a intenciones de búsqueda claramente bibliográficas (*discovery tools* de las bibliotecas, bases de datos bibliográficas, buscadores académicos, ...)? ¿Cómo se puede hacer?
- ¿Qué peso debe tener la medición de la audiencia en la gestión de una revista académica de acceso abierto?
- ¿Es sostenible la creación de contenido editorial en abierto si no se dispone de un número razonable de visitas?

## Referencias

**Aggarwal, Pranjal; Murahari, Vishvak; Rajpurohit, Tanmay; Kalyan, Ashwin; Narasimhan, Karthik; Deshpande, Ameet** (2024). “GEO: Generative Engine Optimization”. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5-16.

<https://doi.org/10.1145/3637528.3671900>

**Bains, Callum** (2024). “The chatbot optimisation game: Can we trust AI web searches?”. *The Guardian*, 3 noviembre.

<https://www.theguardian.com/technology/2024/nov/03/the-chatbot-optimisation-game-can-we-trust-ai-web-searches>

**Bergstrom, Tracy; Ruediger, Dylan** (2024). *A third transformation? Generative AI and scholarly publishing*. Ithaca S+R.

<https://doi.org/10.18665/sr.321519>

**Connaway, Lynn-Sillipigni; Dickey, Timothy J.; Radford, Marie L.** (2011). “‘If it is too inconvenient I’m not going after it’: Convenience as a critical factor in information-seeking behaviors”. *Library & information science research*, v. 33, n. 3, pp. 179-190.

<https://doi.org/10.1016/j.lisr.2010.12.002>

**Cruz-Argudo, Francisco; García-Varea, Ismael; Martínez-Carrascal, Juan A.; Ruiz-Martínez, Antonio; Ruiz-Martínez, Pedro M.; Sánchez-Campos, Alberto; Turró-Ribalta, Carlos** (2024). *La inteligencia artificial generativa en la docencia universitaria: Oportunidades, desafíos y recomendaciones*. CRUE.

[https://www.crue.org/wp-content/uploads/2024/03/Crue-Digitalizacion\\_IA-Generativa.pdf](https://www.crue.org/wp-content/uploads/2024/03/Crue-Digitalizacion_IA-Generativa.pdf)

**Deschenes, Amy; McMahon, Meg** (2024). “A survey on student use of generative AI chatbots for academic research”. *Evidence based library and information practice*, v. 19, n. 2.

<https://doi.org/10.18438/eblip30512>

**Gorichanaz, Tim** (2024). “Virtuous search: A framework for intellectual virtue in online search”. *Journal of the Association for Information Science and Technology*, v. 75, n. 5, pp. 538-549.

<https://doi.org/10.1002/asi.24832>

**Goyanes, Manuel; Lopezosa, Carlos** (2024). “ChatGPT en ciencias sociales: Revisión de la literatura sobre el uso de inteligencia artificial (IA) de *OpenAI* en investigación cualitativa y cuantitativa”. *Anuario ThinkEPI*, v. 18, e18e04.

<https://doi.org/10.3145/thinkepi.2024.e18a04>

**Gordillo-Bernárdez, Saül** (2024). “Àlex Hinojo Sánchez. Director de projectes digitals i activista del coneixement lliure. Viquipèdia davant la IA: Entre la resistència i la irrellevància”. *El Punt Avui*, 11 noviembre.

<https://www.elpuntavui.cat/societat/article/15-ciencia/2477475-viquipedia-davant-la-ia-entre-la-resistencia-i-la-irrellevancia.html>

**Kumar, Aounon; Lakkaraju, Himabindu** (2024). *Manipulating large language models to increase product visibility*. arXiv (No. arXiv:2404.07981).

<http://arxiv.org/abs/2404.07981>

**Li, Xiaoxi; Jin, Jiajie; Zhou, Yujia; Zhang, Yuyao; Zhang, Peitian; Zhu, Yutao; Dou, Zhicheng** (2024). *From matching to generation: A survey on generative information retrieval*. arXiv (No. arXiv:2404.14851).

<https://doi.org/10.48550/arXiv.2404.14851>

**Lopezosa, Carlos; Goyanes, Manuel** (2024). "Evaluación del uso ético de ChatGPT en investigación científica: Principales tendencias, mejores prácticas y desafíos identificados". *Anuario ThinkEPI*, v. 18, e18e19.

<https://doi.org/10.3145/thinkepi.2024.e18a19>

**Messerli, Lisa; Crockett, M. J.** (2024). "Artificial intelligence and illusions of understanding in scientific research". *Nature*, v. 627, pp. 49-58.

<https://doi.org/10.1038/s41586-024-07146-0>

**Miao, Fengchun; Holmes, Wayne** (2023). *Guidance for generative AI in education and research*. Unesco.

<https://doi.org/10.54675/EWZM9535>

**Nafria, Ismael; Nafria, Clara** (2024). "OpenAI lanza ChatGPT Search, su buscador web con IA". *Hoy en internet*, 17 junio.

<https://hoyeninternet.beehiiv.com/p/openai-lanza-chatgpt-search-su-buscador-web-con-ia>

**Potthast, Martin; Hagen, Matthias; Stein, Benno** (2020). "The dilemma of the direct answer". *ACM SIGIR Forum*, v. 54, n. 1, pp. 1-12.

<https://doi.org/10.1145/3451964.3451978>

**Smith, Brad** (2024). "Google AI Overviews are an evolution, not a revolution". *Search engine land*, 25 junio.

<https://searchengineland.com/google-ai-overviews-evolution-revolution-443454>

**Strzelecki, Artur; Rutecka, Paulina** (2020). "Direct answers in Google search results". *IEEE Access*, v. 8, 103642-103654.

<https://doi.org/10.1109/ACCESS.2020.2999160>

**Via, Anna** (2023). "From ancient oracles to Generative AI. A story about human's desire to predict and influence the future". *Medium*, 3 diciembre.

<https://medium.com/artificial-corner/from-ancient-oracles-to-generative-ai-da86f64e0494>

**Vitela-Caraveo, Alex; Urbano, Cristóbal** (2020). "Web analytics for open access academic journals: Justification, planning and implementation". *BiD: Textos Universitaris de Biblioteconomia i Documentació*, v. 45, n. 8.

<https://doi.org/10.1344/BiD2020.45.20>

**Withorn, Tessa** (2023). "Google SGE: A new way to search, teach, and resist". *College & research libraries news*, v. 84, n. 10.

<https://doi.org/10.5860/crln.84.10.373>



## Sobre la posible reducción de visitas y las nuevas técnicas de optimización necesarias

**Rubén Alcaraz-Martínez**



Esta réplica busca, más que ofrecer respuestas, aportar algunos elementos adicionales que puede ser interesante tener en consideración para plantear nuevas preguntas, así como recoger algunos movimientos recientes dentro del mercado tecnológico de la IA que podrían influir aún más en esa transformación en el comportamiento informacional que Cristóbal antoja en su nota. Para ello toma en consideración dos de las preguntas planteadas en la nota original, concretamente las referidas a la posible reducción de visitas y a las nuevas técnicas de optimización que serían necesarias en el caso de buscar mejorar la visibilidad de los contenidos de un dominio entre las respuestas ofrecidas por la IA generativa. Se aleja, no obstante, del ámbito de los portales de revistas académicas.

La consultora *Gartner* predijo hace unos meses una reducción en torno al 25% del volumen de tráfico que manejan los motores de búsqueda en favor de los *chatbots* basados en IA (*Gartner, 2024*). Es decir, consultas que anteriormente se habían resuelto mediante un buscador que ofrecía una serie de resultados entre los cuales había una o más páginas que resolvían la necesidad informacional, empiezan y acaban ahora en una conversación de chat. Si bien los buscadores ya habían incorporado diferentes módulos con respuestas directas en sus páginas de resultados (respuestas directas, fragmentos destacados, bloques de preguntas y respuestas, *knowledge panels* o el más reciente *AI overview*, entre otros), impactando notablemente algunos de ellos en ciertos tipos de sitios web (sitios dedicados a la recopilación de letras de canciones, calendarios con eventos deportivos, resultados de eventos de todo tipo en directo, o incluso afectando a sitios muy populares de preguntas y respuestas como *Quora, Stack Overflow* o *Reddit*) y dando lugar a las conocidas como búsquedas con cero clics (*zero-click search*) (*Silva, 2024*), este cambio de paradigma se acrecienta con la aparición de los *chatbots*, mucho más versátiles y capaces de resolver muchas más intenciones de búsqueda.

Pero ¿qué tipo de consultas o intenciones de búsqueda son las que realmente pueden virar hacia las IA generativas? Las consultas de tipo informacional son las que se antojan más “sustituibles”, pero integraciones tecnológicas recientes como la funcionalidad de compra que *Perplexity* ha lanzado en Estados Unidos (*Perplexity, 2024*), compiten directamente con sistemas que nos acompañan desde hace varios años como *Google Shopping*, al tiempo que ponen en duda la perdurabilidad de estrategias como las telarañas de sitios web centrados en estrategias de marketing de afiliados en las que *Amazon* es una de las empresas referentes.

A medio camino entre las búsquedas informacionales y las transaccionales, las denominadas búsquedas de tipo “investigación comercial”, es decir, las que se centran en analizar el mercado en busca de argumentos para la adquisición de un producto o servicio, también son candidatas, compitiendo directamente con los muy populares comparadores de precios. También lo son las derivadas de la intención de búsqueda “visitar en persona”, relacionadas con el interés en visitar un establecimiento físico, ya que con una conexión en tiempo real y acceso a datos de establecimientos locales, tecnológicamente no resulta demasiado complejo mostrar una lista de sitios cercanos a la posición del usuario ordenados, ahora, no de acuerdo con los criterios del buscador, sino de acuerdo con ciertos atributos que incluso el usuario podría ponderar en su petición (cercanía, valoraciones, horario de cierre, amigable para celíacos...), priorizando u otorgando pesos diferentes a los que los buscadores han implementado a través de sus algoritmos. Finalmente, las intenciones de búsqueda navegacionales (o *go to website*), parecen tener menos sentido que sean asumidas por estas tecnologías.

Cristóbal planteaba la duda sobre cómo posicionar contenidos en este contexto. Lo que podemos intuir es que, si bien se trata de tecnologías diferentes, las bases fundamentales que rigen el SEO, tanto algunas de las más clásicas relacionadas con la optimización del contenido, como con su apartado técnico, seguirán siendo relevantes. También lo serán factores incorporados más recientemente como, por ejemplo, los que emanan de las directrices EEAT (*Experience, Expertise, Authoritativeness, and Trust*) de *Google* (2024). Es decir, desde un punto de vista técnico, será necesario que los sitios web sean compatibles, rastreables y procesables por las IA (SEO técnico); por lo que respecta al contenido, este deberá ser igualmente relevante para la consulta formulada (densidad y ubicación de la palabra clave, etc.) (SEO de contenidos); finalmente, respecto a la calidad de los contenidos, la autoridad del dominio, los enlaces entrantes recibidos, la disponibilidad de información sobre los autores del contenido, las fuentes de referencia utilizadas, entre otros (algunos de ellos parte fundamental de las EEAT), también se antojan como elementos que las IA tomarán en consideración, en tanto que modelos ya implementados como *RankBrain* de *Google*, ya lo hacen. Quizá una de las grandes diferencias radica en la capacidad para manejar, procesar y sintetizar contenidos que pre-

senta la IA sin necesidad de que estos estén altamente estructurados mediante marcas (etiquetas h1-h6, meta títulos, datos estructurados con sintaxis Schema.org...).

Otro elemento que resultará importante tener en cuenta es qué *chatbots* utiliza la audiencia a la que se dirige el contenido o producto. Salvando las distancias, al hacer SEO, si nos dirigimos al mercado español, nos interesa trabajar la visibilidad en *Google* porque es, de lejos, el más utilizado en el territorio, pero si nos dirigimos a otros mercados, por ejemplo, al ruso, son otros buscadores como *Yandex*, donde interesa aparecer. En este sentido, si *ChatGPT* o *Copilot* utilizan *Bing*, la visibilidad en esa plataforma dependerá en gran medida del trabajo SEO realizado en ese buscador. La misma *Microsoft* publicó hace unos días un artículo en el blog de *Bing* sobre el valor del análisis de la intención de búsqueda en las consultas potenciadas con IA (*Microsoft*, 2024), donde destacaba la importancia de:

- Trabajar más para satisfacer la intención de búsqueda y menos la relevancia para unas palabras clave determinadas (**Alcaraz-Martínez**, 2021).
- Optimizar el contenido para palabras clave de cola larga (muy específicas).
- Incluir palabras clave conversacionales, es decir, consultas que se redactan siguiendo patrones similares a los utilizados cuando hablamos con otras personas.
- Relacionado con el punto anterior, integrar palabras clave en forma de pregunta, a las cuales el mismo contenido debería dar respuesta.
- Potenciar los textos con palabras clave relacionadas con la palabra clave principal. Algo que en el SEO se conoce como LSI (*Latent Semantic Indexing*) *keywords*.

De la misma manera que en los últimos años los profesionales del SEO han trabajado la visibilidad de sus contenidos en los fragmentos destacados de la *SERP* de *Google*, parece que la optimización en un entorno de chat pasará, al menos inicialmente, por estrategias similares. En este sentido, lo que se ha trabajado por el momento en el sector para adaptarse a las búsquedas de clic cero en las *SERP* es competir por palabras clave de tipo *long tail* (muy específicas) capaces de disparar fragmentos destacados de diferente tipo (algo que se puede averiguar con herramientas como *Semrush* o simplemente realizando búsquedas) para asegurar una posición relevante de la marca o dominio en esos espacios destacados, marcando con encabezados (h2-h3) las preguntas a las que responde el contenido y aportando una respuesta breve a continuación, mejorando la lecturabilidad del contenido (frases cortas, párrafos breves, uso de listas...), entre otras formas de optimización.

Para finalizar, algunas preguntas que surgen a partir del texto de Cristóbal y de esta discreta aportación podrían ser:

- ¿Cómo impactan estos cambios en la diversidad de fuentes de información? ¿Es decir, el predominio de los *chatbots* basados en IA puede centralizar aún más la información en unas pocas fuentes, limitando la diversidad al depender estas herramientas de un número limitado de recursos confiables y compatibles?
- ¿Cómo afectará al contenido? ¿Será algo beneficioso para los usuarios (contenido de mayor calidad) o, por el contrario, lo que se impondrá es un contenido más sobreoptimizado y antinatural del que ya lo es hoy? ¿Se impondrá un modelo de contenido basado en bloques modulares fácilmente reutilizables por las IA? ¿Tecnologías como Schema.org seguirán teniendo sentido?
- ¿Integrarán publicidad –resultados patrocinados– a corto, medio o largo plazo?, siguiendo un camino paralelo al que ya exploraron los buscadores?

## Referencias

**Alcaraz-Martínez, Rubén** (2021). "SEO centrado en el usuario: un cambio de paradigma orientado a satisfacer mejor las necesidades de las personas en el buscador de Google". *Cuadernos de documentación multimedia*, v. 32. <https://doi.org/10.5209/cdmu.81156>

Gartner (2024). "Gartner predicts search engine volume will drop 25% by 2026, due to AI chatbots and other virtual agents". Press release. Newsroom.

<https://www.gartner.com/en/newsroom/press-releases/2024-02-19-gartner-predicts-search-engine-volume-will-drop-25-percent-by-2026-due-to-ai-chatbots-and-other-virtual-agents>

Google (2024). Search quality evaluator guidelines.

<https://static.googleusercontent.com/media/googleusercontent.com/media/guidelines.raterhub.com/ca//searchqualityevaluatorguidelines.pdf>

Microsoft (2024). "The value of intent-driven SEO in AI-powered search engines". Microsoft Bing blogs.

<https://blogs.bing.com/webmaster/November-2024/The-Value-of-Intent>

Perplexity (2024). Shop without limits.

<https://www.perplexity.ai/shopping>

**Silva, Carlos** (2024). "What are zero-click searches & how do they impact SEO?". *Semrush blog*, 10 mayo.

<https://www.semrush.com/blog/zero-click-searches>

**Rubén Alcaraz Martínez**

Universitat de Barcelona

[ralcaraz@ub.edu](mailto:ralcaraz@ub.edu)